Optimal Replacement in the Proportional Hazards Model with Semi-Markovian Covariate Process and Continuous Monitoring

Xiang Wu[1], and Sarah M. Ryan[2], IEEE member

ABSTRACT

Motivated by the increasing use of condition monitoring technology for electrical transformers, this paper deals with the optimal replacement of a system having a hazard function that follows the proportional hazards model with a semi-Markovian covariate process, which we assume is under continuous monitoring. Although the optimality of a threshold replacement policy to minimize the long-run average cost per unit time was 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 is 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. Numerical examples and sensitivity analysis provide some insights into the suitability of a Markov approximation, and the sources of variability in the cost. The optimization method developed here is much more efficient than the approach that approximates continuous monitoring as periodic, and then optimizes the periodic monitoring parameters.

[1] Industrial and Manufacturing Systems Engineering, Iowa State University, Ames, IA, USA, xiangwu@iastate.edu

[2] Industrial and Manufacturing Systems Engineering, Iowa State University, Ames, IA, USA, smryan@iasate.edu

1

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

ACRONYMS

CBM   Condition-based maintenance

DGA   Dissolved gas analysis

PH   Proportional hazards

STD   Sojourn time distributions

CV   Coefficient of variation

NOTATION

t The age of the current system.

$Z = \{Z_t, t \geq 0\}$ A right continuous semi-Markov process with a finite state space

$\{0, 1, ..., n-1\}$ and $Z_0 = 0$ that reflects the health condition of the system

at age t.

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

$\psi(Z_t)$ The link function in PH model that depends on the state of the covariate

process Z.

X_k The sojourn time of the Z process in state k, $k = 0, ..., n-2$.

$f_{X_k}(x_k)$ The pdf of X_k, $k = 0, ..., n-2$.

S_k The age at which the covariate state changes from k to $k+1$,

$$k = 0, ..., n-2 \, .$$

$g_k(s_0, s_1, ..., s_k)$ 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_{T_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, $K > 0$.

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 $\psi(0) = 1$.

4. The practice of continuous monitoring influences neither the covariate process Z

    nor the system failure process.

1   INTRODUCTION

This article concerns a condition-based maintenance (CBM) problem for critical assets.

Compared to classical preventive maintenance, CBM improves the decision-making process

by exploiting available information about the system's operating conditions. Increasingly,

condition monitoring technology is gaining favor as a way to diagnose the health status, and detect the impending failure of expensive assets.

This work was motivated by the need to improve the management of capital-intensive assets such as high-voltage power transformers. As explained by Wang et al. [1], "As transformers age, their internal condition 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." Unexpected failure of power transformers results in unscheduled outages with power delivery problems, and may cause immense economic loss. For example, the replacement cost of a single phase 500 MVA transformer is around 1 million dollars, while the failure cost could run several times as high as that number [2]. To reduce the risk of unexpected failure, on-line monitoring has become common practice, and the condition information concerning transformers in the field can be returned in real time to a central location for continuous assessment [2], [3]. A real example is described in [2], where on-line dissolved gas analyzers attached to transformers collect dissolved gas analysis (DGA) data six times each day, on a regular 4-hour schedule. In view of the multi-decade life cycle of transformers, it is adequate to view this kind of practice as continuous monitoring. The high cost of unexpected transformer failure motivates our study of how to make best use of the condition information to decide when to perform preventive replacement.

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, and extensive research has been done with Markov and semi-Markov decision models [4] - [9]. Douer and Yechiali [6] studied the optimal repair and replacement problem in Markovian systems, and they introduced a generalized control limit policy which is optimal under reasonable conditions. Lam and Yeh [7] used a semi-Markov process to model a multi-state deteriorating system, and considered state-age-dependent replacement policies. They showed that optimal replacement policies have monotonic properties under reasonable assumptions on replacement cost, replacement time, and failure rate. Chen and Trivedi [8] built a semi-Markov decision model for condition-based maintenance policy optimization, and presented an approach to optimize the inspection rate and maintenance policy jointly. The issues of imperfect monitoring in state-based preventive maintenance were considered in [10], [11]. In contrast to the multi-state deteriorating models, Toscano and Lyonnet [12] 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 way to incorporate condition information into risk estimation is the proportional hazards (PH) model [13], which explicitly includes both the age and the condition information in the calculation of the hazard function. It combines a baseline hazard function which accounts for the aging degradation with a link function that takes the condition information into account to improve the prediction of failure. Generally, the condition information is described by a multi-state covariate (diagnostic)

process $Z = \{Z_t, t \geq 0\}$. The PH-based replacement policies have been successfully applied in a variety of industrial sectors such as pulp and water, coal plants, nuclear plant refueling, military land armored vehicles, construction industry backhoes, marine diesel engines, and turbines in a nuclear plant [14].

Several papers have been published to optimize the decision-making in the PH model setting. Makis and Jardine [15] investigated the optimal replacement policy for systems under a PH model 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 [16] extended Makis and Jardine's model by relaxing the monotonicity assumption of the hazard function, and they developed methods for parameter estimation in the PH model as well. The same model was extended in [17] 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 [18] removed the discrete-time approximation of the continuous time covariate process in [15], 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 whether the investment of condition monitoring technology in capital-intensive physical assets is worthwhile. All of these papers assumed the covariate processes to be Markov processes, and under periodic monitoring.

In this paper, we extend the PH-based replacement models to systems with semi-Markov covariate processes under continuous monitoring. We consider parametric PH model with a baseline hazard function, and a time dependent covariate process. In the transformer

application, it is reasonable to let the covariate process Z represent the condition of the insulation, which degrades over time, and may be classified into several different states, such as new, normal, warning, and dangerous. Assume the state of the insulation can be perfectly inferred from a combination of monitored variables including 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]. By modeling the evolution of the insulation state, the hazard function for the transformer can be evaluated, and further, the mean time to failure and the average cost associated with any given replacement policy can be calculated. In the PH model setting, a transformer failure can occur from any insulation state with increasing risk of failure as the insulation condition degrades.

Maintenance to improve the condition of the insulation requires taking the transformer out of service for a significant period of time to replace the insulation, which is not a practical option. Besides, the maintenance cost is relatively low compared to the preventive replacement cost plus the failure cost [2], [19]. Thus, in this paper, we consider replacement of the transformer as the only maintenance option.

Examining existing PH-based replacement models exposed the gaps between the literature and practice. So far, the form of the optimal policy for systems under continuous monitoring has not been articulated, and how to estimate the risk with continuously monitored information has not been addressed. In addition, a Markovian model may not be appropriate for the covariate process. Requiring that times between transitions among the covariate states be exponentially distributed is an added approximation which limits the usage of the model. Therefore, we adopt a semi-Markov covariate process with general transition

time distributions.

The contributions of this research and outline of the paper are as follows. By identifying our model as a special case of the one described in [20], 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 use conditioning arguments to derive explicit expressions for *s*-expected life and failure probability of transformers in terms of the policy parameters in Section 4. The iterative procedure developed by Bergman [20] is specified for our model to find the optimal threshold ages. The model and the solution procedure are illustrated by numerical examples in Section 5. We discuss its computational advantage over the recursive procedure [18], and we study the effect of different sojourn time distributions of the covariate process on the optimal policy and cost. In addition, sensitivity analysis is performed on a specific instance to demonstrate how the variations in the input parameters would affect the long-run average cost.

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 value of covariates that reflect the health condition of the system.

For simplicity, we consider only one covariate. To account for both the age effect and the

8

condition information in the system's hazard function, the PH model 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

$$h(t, Z_t) = h_0(t)\psi(Z_t), t \geq 0.$$ (1)

We assume that $Z = \{Z_t, t \geq 0\}$ is a continuous-time semi-Markov process which depicts the evolution of the covariate, and is under continuous monitoring. It has a finite state space $\{0, 1, ..., n-1\}$, where state 0 represents the covariate state corresponding to a new system, and states $1, 2, ..., n-1$ reflect the increasingly deteriorating condition. It follows that the conditional survivor function is given by

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

From this function, we can see that a system failure can occur in any state at any time with increasing likelihood as the system ages, and the health condition degrades.

Between any two consecutive replacements, the covariate process Z changes states according to a pure birth process; i.e., whenever a transition occurs, the state of the process always increases by one, and state $n-1$ is absorbing. Replacement is instantaneous, and the covariate returns to state 0 upon replacement. The time interval between two successive transitions is a random variable with any distribution. Let X_k be the sojourn time in state k. We allow X_k to follow an arbitrary distribution with density $f_{X_k}(x_k)$, for $k \leq n-2$; the distribution of X_{n-1} is immaterial because the covariate process exits from that state only when the system is replaced. Define $S_k = \sum_{i=0}^{k} X_i, k = 0, 1, .., n-2$, which is the age when the covariate moves from state k to state $k+1$. The joint pdf, and Cdf of $S_0, S_1, ..., S_k$, for $k = 0, ..., n-2$, are represented as $g_k(s_0, s_1, ..., s_k)$, and $G_k(s_0, s_1, ..., s_k)$ respectively, where

$0 < s_0 < s_1 < ... < s_k$. As will be shown in Section 4, the pdf $g_k(s_0, s_1, ..., s_k)$ is fully determined by $f_{X_k}(x_k)$, $k = 0, ..., n-2$, as is $G_k(s_0, s_1, ..., s_k)$.

In practice, the state of the covariate is inferred from continuously monitored variables. In the transformer application, the state of insulation is determined by a combination of acoustic and electrical signals, detection of moisture or gases in the insulating oil, dissolved gas analysis data, and so on. By carefully examining historical data, the point in time at which the covariate changes state would be known, and the forms and the parameters of $f_{X_k}(x_k)$ could be identified and estimated using standard statistical methods.

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.

Define the replacement rule δ_{T_d}: Replace at failure or at T_d, whichever occurs first. Utilizing the classical cost structure, assume each planned replacement costs $C > 0$, and each failure replacement incurs an additional cost $K > 0$. Then, according to the theory of renewal reward processes [21], the long run average cost per unit time can be expressed as

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

where $\Pr(T_d \geq T)$ is the probability of failure replacement, and $E[\min\{T, T_d\}]$ is the *s*-expected replacement time. The main 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; and to establish procedures to obtain the parameters of the optimal policy.

10

3 THE FORM OF THE OPTIMAL REPLACEMENT POLICIES

Bergman [20] investigated the optimal replacement problem under a general failure model, in which the hazard rate $h(\cdot)$ of system failure is non-decreasing, and completely determined by a general stochastic process $X(t)$, $t \geq 0$. It is assumed that $X(t)$ is also non-decreasing, and under continuous monitoring. Under the same cost structure as in Section 2, Bergman showed that the optimal replacement policy is of the control limit type, and the optimal stopping time has the form

$$T_d^* = \inf\{t \geq 0 : h(X(t)) \geq d^* / K\} \tag{4}$$

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.

Equation (4) indicates that, for a given control limit d^* / K, the optimal policy parameters can be calculated. However, d^* itself is dependent on the optimal policy. To solve this difficulty, Bergman proved the following proposition, which leads to an iterative algorithm that produces a sequence converging to an optimal cost.

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

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

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

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

A generalization made in the latter part of [20] greatly extends the application scope of this model. Therein Bergman stated that the process $X(t)$ can be generalized to be a stochastic vector process with $X = (X_1, X_2, \ldots, X_n)$, which represents n different

measurements of deterioration. As long as each component of $X(t)$ is non-decreasing, and the state-dependent hazard rate function $h\big(X(t)\big)$ is non-decreasing in each component of $X(t)$, the above conclusions hold.

The PH model with a 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 Z_t as the other component of $X(t)$. Thus we obtain the following theorem.

Theorem 1. *For a system whose failure time follows the proportional hazards model* (1) *that is to be replaced at the smaller of its failure time or a replacement stopping time, the optimal stopping time satisfies*

$$T_d^* = \inf\{t \geq 0, h_0(t)\psi(Z_t) \geq d^* / K\} \tag{7}$$

where d^* *is the optimal cost.*

The optimal replacement policy specified by (7) may be explained as: *replace at failure or when the hazard rate of the system reaches or exceeds a certain level (control limit).* 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, (7) determines a unique threshold age for replacement because the hazard rate function is monotonic in time. Hence, the optimal replacement policy for our model can be uniquely defined by n threshold ages. Consider a system with a three-state Z process. As illustrated in Fig. 1, the control limit d^* / K for the hazard rate fixes the planned replacement ages t_0, t_1, t_2 for state 0, 1, 2 respectively. Because the link function increases with the covariate
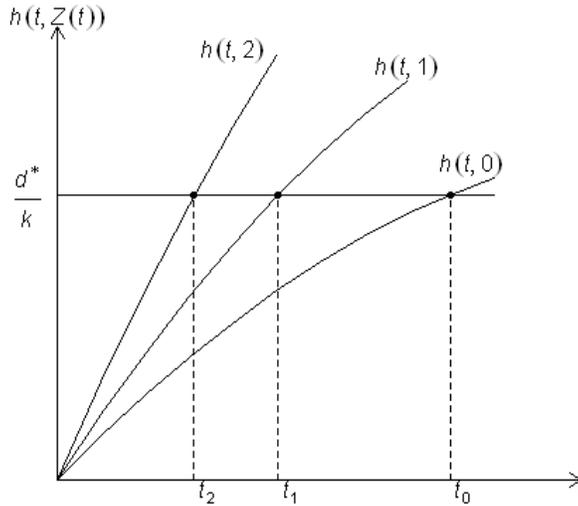
state, we have $t_0 > t_1 > t_2$.

Fig. 1. Replacement ages defined by the control limit.

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

With the form of the optimal policy known from Theorem 1, and the iterative algorithm given in Proposition 1, there is still one barrier in the way of obtaining the optimal policy and cost for our model, which is the evaluation of (6), or how to compute the corresponding cost for a given stopping rule. An explicit expression for the objective function (3) in terms of the policy parameters $t_0, t_1, ..., t_{n-1}$ is necessary to overcome the barrier. We address this issue in the next section.

4   EXPLICIT EXPRESSION OF THE LONG-RUN AVERAGE COST

From (3), calculation of the objective involves evaluating the failure probability

13

$\Pr(T_d \geq T)$, and s-expected time to replacement $E\big[\min\{T,T_d\}\big]$. For notational convenience,

define $W_d = W(t_0, t_1, ..., t_{n-1}) = E\big[\min\{T, T_d\}\big]$ as the s-expected life of the system, and define

$Q_d = Q(t_0, t_1, ..., t_{n-1}) = \Pr\big(T \leq T_d\big)$ as the probability of failure under policy

$\delta_{T_d} = \{t_0, t_1, ..., t_{n-1}\}$. In what follows, we show that it is possible to explicitly represent W_d

and Q_d as functions of $t_0, t_1, ..., t_{n-1}$ by conditioning on the time instants at which the system

changes state; 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 results generalize to situations with more states.

Assume the marginal pdfs of sojourn times X_0, and X_1 are $f_{X_0}(\cdot)$, and $f_{X_1}(\cdot)$

respectively. It follows that the pdf of S_0 is

$$g_0(s_0) = f_{X_0}(s_0). \tag{8}$$

Also, note that the event $\big[S_0 = s_0, S_1 = s_1\big]$ is equivalent to the event $\big[X_0 = s_0, X_1 = s_1 - s_0\big]$.

Hence the joint pdf of S_0 and S_1 is

$$g_1(s_0, s_1) = f_{X_0}(s_0) f_{X_1}(s_1 - s_0), \quad 0 < s_0 < s_1. \tag{9}$$

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 $F(t; s_0, s_1) \equiv \Pr(T \leq t \mid S_0 = s_0, S_1 = s_1)$.

Then, for $t \leq s_0$,

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

For $s_1 > t > s_0$,

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

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 discussed below. Note that $t_2 < t_1 < t_0$, and $s_0 < s_1$. Under each case, the expressions of W_d and Q_d can be derived accordingly.

Let

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

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

By the Law of Iterated Expectation [22],

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

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

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

$$W(t_0, t_1, t_2; s_0, s_1) = W_2(s_0, t_1) \equiv \int_0^{s_0} t\,dF_0(t) + \int_{s_0}^{t_1} t\,dF_1(s_0, t) + t_1\left[1 - F_1(s_0, t_1)\right]$$

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

15

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

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

$$W(t_0, t_1, t_2; s_0, s_1) = W_3(s_0, s_1) \equiv \int_0^{s_0} t\,dF_0(t) + \int_{s_0}^{s_1} t\,dF_1(s_0, t) + s_1\left[1 - F_1(s_0, s_1)\right]$$

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

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

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

$$W(t_0, t_1, t_2; s_0, s_1) = W_4(s_0, s_1, t_2) \equiv \int_0^{s_0} t\,dF_0(t) + \int_{s_0}^{s_1} t\,dF_1(s_0, t) + \int_{s_1}^{t_2} t\,dF_2(s_0, s_1, t) + t_2\left[1 - F_2(s_0, s_1, t_2)\right]$$

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

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

$$W_d = E\left[E\left(\min\{T, T_d\} \mid S_0, S_1\right)\right] = \int_{t_0}^{\infty} W_0(t_0) g_0(s_0)\,ds_0 + \int_{t_1}^{t_0} W_1(s_0) g_0(s_0)\,ds_0 \tag{10}$$
$$+ \int_{t_1}^{\infty} \int_0^{t_1} W_2(s_0, t_1) g_1(s_0, s_1)\,ds_0\,ds_1 + \int_{t_2}^{t_1} \int_0^{s_1} W_3(s_0, s_1) g_1(s_0, s_1)\,ds_0\,ds_1 + \int_0^{t_2} \int_0^{s_1} W_4(s_0, s_1, t_2) g_1(s_0, s_1)\,ds_0\,ds_1$$

$$Q_d = E\left[P(T \le T_d \mid S_0, S_1)\right] = \int_{t_0}^{\infty} Q_0(t_0) g_0(s_0)\,ds_0 + \int_{t_1}^{t_0} Q_1(s_0) g_0(s_0)\,ds_0$$
$$+ \int_{t_1}^{\infty} \int_0^{t_1} Q_2(s_0, t_1) g_1(s_0, s_1)\,ds_0\,ds_1 + \int_{t_2}^{t_1} \int_0^{s_1} Q_3(s_0, s_1) g_1(s_0, s_1)\,ds_0\,ds_1 + \int_0^{t_2} \int_0^{s_1} Q_4(s_0, s_1, t_2) g_1(s_0, s_1)\,ds_0\,ds_1$$

(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 the system with a three-state covariate process. For a system with an n-state covariate process, 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. Explicitly writing out the $(2n - 1)$ n-fold integrals seems to be a formidable task. However, thanks to the connection between the n state model and the $(n + 1)$ state model,

16

this task is reduced to something tractable. In fact, for the $(n+1)$ state model, the expression

for 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 model, and the last three cases of which form a partition of

the last case of the n state model by values of the new transition instant, S_n. Therefore, we

can build the expressions of W_d and Q_d for an n-state covariate process by adding one

state at a time. For comparison and illustration, we show the formulas for a system with a

two-state covariate process in the APPENDIX.

Based on the explicit expressions of W_d, Q_d, and Proposition 1, we describe the

following iterative algorithm, which can 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.,

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

3. Use the replacement policy $\delta_m = \{t_0^m, t_1^m, .., t_{n-1}^m\}$ obtained in step 2, (3), (10), and (11)

   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, we consider a system with a three-state covariate process as a numerical example. In the following analysis, we assume that the functions that define the failure model, namely $h_0(t)$, $\psi(Z_t)$, and $f_{X_k}(x_k)$, are known, and their parameters are given (estimated). In practice, with historical monitoring data and lifetime data, the forms of those functions can be established either empirically, or through careful statistical analysis [23], [24]. The parameters of those functions can be estimated using the maximum likelihood method and its variants (to cope with the truncated and censored data), such as the one used in [16].

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 $K = 25$. Because the forms of $h_0(t)$ and $\psi(Z_t)$ are predefined, the PH model here is parametric rather than semi-parametric, as described in [13].

Suppose the semi-Markov process Z has three states $\{0, 1, 2\}$, and the sojourn times X_0 and X_1 are s-independent identically distributed Weibull random variables with mean 1. The Weibull distribution is chosen here because 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

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

with $\beta = 1.5$, and $\eta = 1.1077$. It is not hard to check that the mean of X_i is approximately

18

1.

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 (10) 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 I. The iterative algorithm converges after five iterations to the optimal average cost $d^* = 23.4364$. The algorithm was implemented in *Mathematica®*.

Table I

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 II 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 II 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 distributions, 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 higher replacement cost. The cost errors for using policy parameters from the Markov model in other sojourn times are shown in Table III. We can see that the relative error becomes smaller as the CV of the true STD gets closer to 1. In this example, those errors are relatively small, which means that, 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 of only 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.

In the special case where the covariate process is Markovian; i.e., the STD is Weibull(1, 1), the computational procedure for periodic monitoring [18] can be used to approximate continuous monitoring by setting the monitoring interval to be very small. In that approach, recursion is needed for calculation of both $W(t_0, t_1, t_2)$ (s-expected life), and $Q(t_0, t_1, t_2)$ (failure probability), while the approach derived in this paper requires no recursion. This result gives the current approach a great computational advantage. For the example discussed here with a Weibull(1, 1) STD, the computational time to obtain the optimal policy and cost is 0.25 seconds on a computer with 1.83 GHz CPU, and 2GB main memory. However, if using periodic monitoring with an interval of 0.01 time units to approximate the continuous

monitoring, the resulting policy is similar, but the computational time is 10.4 seconds, which is substantially longer. Based on this computational advantage, we suggest using the formulas in this paper to approximate the optimal policy under periodic monitoring when the monitoring interval is small, as well as to compute the exact optimal policy under continuous monitoring.

Table II

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 III

Cost errors for using policy parameters from a 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 IV shows the optimal policies and costs for Lognormal STDs. Again, all of these

21

distributions have the same mean, approximately equal to 1. Table VI confirms the conclusion that a large CV for the STD has a harmful effect on the optimal cost. Besides, comparing similar cases in Table II and Table IV suggests that, for the same CV, the Lognormal sojourn time leads to a lower optimal cost than the Weibull sojourn time.

Table IV

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 fixed. However, in practice, some of those parameters must be estimated from the historical data of the system. The quality of the estimates will directly affect the validity of the resulting replacement policy. In this subsection, we investigate how the variations in the model parameters impact the long-run average cost, and we assess the relative importance of model parameters through sensitivity analysis. In particular, we evaluate three input parameters, which are *a* and *b* in the baseline hazard function, $h_0(t) = bt^{b-1}/a^b$; 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

22

the STD for the Z process.

Assume the true parameter values are $a = 2, b = 2, c = 2$, and their estimates $\hat{a}$, $\hat{b}$, and $\hat{c}$ each s-independently follow the distribution $N(2, 0.4)$. Performing the FAST sensitivity analysis method [25] with 1000 samples using SimLab [26], we get the FAST first-order indexes, as shown in Table V. 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 V

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. The aging and deterioration process is characterized by 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

23

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 [20], and determined 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 threshold 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 $n=3$ covariate states 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 increase the cost of the optimal replacement policy. 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 indicates 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 uncertainty in the monitoring process, that is, the partial observation problem, to our current model; and 3) use a new failure model to relate the covariate information to system failure time distribution, such as an accelerated failure time

model [23].

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 s-expected life of the system $W_d = W(t_0, t_1) = E\left[\min\{T, T_d\}\right]$, and the probability of failure $Q_d = Q(t_0, t_1) = \Pr\left(T \leq T_d\right)$ under policy $\delta_d = \{t_0, t_1\}$ by conditioning on S_0.

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

$$F(t; s_0) \equiv \Pr(T \leq t \mid S_0 = s_0),$$

where s_0 is the realization of S_0.

Then for $t \leq s_0$,

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

For $t > s_0$,

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

Let

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

$$Q(t_0, t_1; s_0) \equiv \Pr(T \leq T_d \mid S_0 = s_0).$$

By the Law of Iterated Expectation [22],

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

There will be three cases.

Case 0: If $s_0 > t_0$, then

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

25

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

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

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

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

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

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

Case 2: If $s_0 < t_1$, then

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

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

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

Then by another application of the Law of Iterated Expectation,

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

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

Comparison with (10) and (11) shows the recursive nature of these expressions.

REFERENCES

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

[2] B. Augenstein, "Outside experts monitor status of key transformers," *Transmission & Distribution World*, May 2003. http://tdworld.com/mag/power_outside_experts_monitor/ (viewed September 2010).

[3] W. Q. Meeker, "Trends in the statistical assessment of reliability," tech. rep., Iowa State University, 2009.

[4] H. Mine and H. Kawai, "An optimal inspection and replacement policy," *IEEE Transactions on Reliability*, vol. 24, pp. 305–309, 1975.

[5] 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.

[6] N. Douer and U. Yechiali, "Optimal repair and replacement in Markovian systems," *Communications in Statistics – Stochastic Models*, vol. 10, pp. 253–270, 1994.

[7] 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.

[8] 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.

[9] A. Pavitsos and E. G. Kyriakidis, "Markov decision models for the optimal maintenance of a production unit with an upstream buffer," *Comput. Oper. Res.*, vol. 36, no. 6, pp. 1993–2006, 2009.

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

[11] J. Ivy and S. 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.

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

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

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

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

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

[17] 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.

[18] X. Wu and S. M. Ryan, "Value of condition monitoring for optimal replacement in the proportional hazards model with continuous time degradation," *IIE Transactions*, vol. 42, pp. 553–563, 2010.

[19] D. J. Woodcock, "Maintenance strategies for the new millennium – condition appraisal of power transformers," 2000.

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

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

2003.

[22]P. Billingsley, *Probability and Measure*. New York, NY: John Wiley & Sons, Inc., 3rd ed., 1995.

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

[24]Y. Hong, W. Q. Meeker, and J. McCalley, "Prediction of remaining life of power transformers based on left truncated and right censored lifetime data," *Annals of Applied Statistics*, vol. 3, pp. 857–879, 2009.

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

[26]European Union, "Simlab – sensitivity analysis," 2009. http://simlab.jrc.ec.europa.eu/ (viewed September 2010).

Xiang Wu is a Ph.D. candidate in the Department of Industrial and Manufacturing Systems Engineering at Iowa State University. He is currently employed with AOL Inc. He received a M.S. degree in Statistics from Iowa State University. He also obtained a B.S. degree in Automation, and a M.S. degree in Control Science and Engineering from Tsinghua University, Beijing, P. R. China. His research interests include stochastic modeling and optimization, reliability and maintenance, and supply chain.

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.