IIE Transactions (2003) **35**, 1091–1101

ISSN: 0740-817X print / 1545-8830 online
DOI: 10.1080/07408170390237162

Cost-variability-sensitive preventive maintenance considering management risk

YONG CHEN[1] and JIONGHUA JIN[2,*]

[1]*Department of Mechanical and Industrial Engineering, The University of Iowa, Iowa City, IA 52242-1527, USA*
E-mail: yongchen@engineering.uiowa.edu
[2]*Department of Systems and Industrial Engineering, The University of Arizona, P.O. Box 210020, Tucson, AZ 85721-0020, USA*
E-mail: judy@sie.arizona.edu

Received June 2001 and accepted December 2002

Traditional preventive maintenance policies, such as age replacement, periodic replacement under minimal repair, and replacement policy N, are all studied based on the expected cost criteria without considering the management risk due to the cost variability. As a result, these policies could be significantly beyond the anticipated maintenance budget allocation and lead to crisis. In order to solve this problem, a new analysis methodology is proposed in this paper to consider the effects of both cost expectation and cost variability on the optimal maintenance policy. A new concept of the *long-run variance* of the cost is defined to represent the maintenance management risk, and then the objective function is revised accordingly to achieve an optimal cost-variability-sensitive maintenance policy. Based on the proposed framework, three traditional preventive maintenance policies have been reinvestigated and the effect of the variability sensitivity on the optimal policies is further analyzed, which reveals general management insights and explicates the search bound of the optimal solution. An example is given to illustrate the importance and the effectiveness of the proposed methodology. Compared with the traditional optimal maintenance policy, the numerical solution shows that the proposed variability-sensitive optimal policies can significantly reduce the maintenance management risk with only a small increase in the expected cost.

Nomenclature

μ = mean time to failure of an item;
a = preventive replacement interval;
$r(t)$ = failure rate function of an item;
λ = cost-variability-sensitive factor;
$A(N)$ = system availability under replacement policy N;
C_t = cost at time unit t;
c_1 = cost of an unexpected/failure replacement for the age replacement policy;
c_2 = cost of a scheduled/preventive replacement for the age replacement policy;
c_r = replacement cost in periodic replacement under minimal repair and replacement policy N;
c_m = minimal repair cost in periodic replacement under minimal repair and replacement policy N;
$S(a)$ = the mean of the length of the replacement cycle for the age replacement policy.

*Corresponding author

1. Introduction

Most of the preventive maintenance policies in the literature are evaluated with only the expected cost criteria (Barlow and Proschan, 1965; Fox, 1966; Glasser, 1967; Nakagawa and Osaki, 1974; Zhang and Jardine, 1998; Liu *et al.*, 2001). However, it is generally more critical to consider not only the expectation of the maintenance cost, but also the variability of the cost. For example, there are two maintenance policies A and B. Under policy A, the maintenance costs \$10 000 per month. Under policy B, there is a 80% probability of a zero cost and a 20% probability of a \$50 000 cost for each month. For both policies, the long-run average of the costs are the same being \$10 000 per month. If the maintenance budget allocation is \$10 000 per month, that will always be affordable under policy A. However, under policy B, the large variability of the cost from one month to another leads to a high management risk under the anticipated budget allocation. As a result, policy A is preferable to policy B in terms of the budget allocation management because the cost of policy A has less cost variability. The maintenance cost variability is generally quite significant due to the unexpected nature of failures. Reduction of this variability, as well as the expectation, of the cost will significantly benefit from effective and efficient management of maintenance cost.

The well-known approach to deal with the cost-variability-risk features stems from the earlier work of Markowitz (1959) on portfolio analysis, where both the mean and variance of the cost are used to characterize the problem. The idea of using variability-sensitive decision-making is further adopted in the research area of the Markov Decision Processes (MDPs). A detailed survey of the related work in this area can be found in White (1988). Some more recent research results in variability-sensitive MDPs are given in Filar and Lee (1984) and Sobel (1985, 1994).

Although the variability-sensitive MDP has been studied as a general topic, little research literature can be found for the variability-sensitive maintenance policies. Using the discounted cost criterion, Tapiero and Venezia (1979) treated the variance of the cost as a risk factor to study a maintenance problem with a fixed replacement cycle. In Tapiero and Venezia (1979) the machine degradation is modeled with a Markov chain whose transition probabilities are proportional to the maintenance expenditure. Rangan and Grace (1988) employed the variance optimization criterion to obtain the optimal replacement cycle for systems subject to shocks. However, there is no further discussion about the effect of the variability sensitivity on the optimal policy in Rangan and Grace (1988). Moreover, both of these works consider only maintenance policies with replacement cycles of equal lengths. With replacement cycles of equal lengths, the variance of the cost can be simply defined based on the variation of the costs from one replacement cycle to another as in Tapiero and Venezia (1979) and Rangan and Grace (1988). However, the lengths of the replacement cycles may be different for a lot of preventive maintenance policies such as age replacement and replacement policy N. For these policies, the cost variation from one replacement cycle to the next is no longer a reasonable risk index. In order to study the cost-variability-sensitive preventive maintenance policies, a general index of the cost variations from one time unit to the next in a long-run cost stream should first be defined.

In this paper, the cost-variability-sensitive polices are fully investigated for some typical preventive maintenance problems, such as age replacement, periodic replacement under minimal repair, and replacement policy N to achieve optimal maintenance management. The *long-run variance* of the cost suggested by Filar *et al.* (1989) for general MDPs is used in this paper to define the cost variation. After the Introduction, the concept of the *long-run variance* of the cost is introduced in Section 2, where a new cost-variability-sensitive optimization criterion is defined. Sections 3 and 4 provide detailed discussions on the variability-sensitive optimal decisions for various maintenance policies, such as age replacement, periodic replacement under minimal repair, and replacement policy N. Comparison studies between the traditional maintenance policies using the expected cost criterion, which is called the variability-neutral policy in this paper, and the optimal cost-variability-sensitive policies have been conducted. It has been shown that for each of the discussed preventive maintenance problems, the variability-sensitive policy is more conservative. In other words, the variability-sensitive policy tends to conduct more maintenance actions with a lower cost for each action to avoid the risk of the higher cost actions. A numerical example is given in Section 5 to demonstrate the effectiveness and the potential applications of the proposed maintenance policy. Section 6 summarizes the paper and discusses future work.

2. Long-run variance of the cost

In order to study a variability-sensitive policy, the measure of the cost variability needs to be first defined. Several measures of variability for stochastic processes with rewards have been discussed and compared in Filar *et al.* (1989). In this paper, the *long-run variance* measure recommended by Filar *et al.* (1989) will be used. A discrete time scale is considered. Let $t = 1, 2, \ldots$ denote the discrete time units, and $C_t(\pi)$ denote the cost spent at time unit t under maintenance policy π. The time unit can normally be chosen as a common time period such as a day, a week, a month, or a year; and the cost variations among these individual time units are considered by the decision makers. It is reasonable to assume that the average lifetime of items is at least two time units, i.e., $\mu \geq 2$; and the preventive replacement time interval is at least one time unit, i.e., $a \geq 1$.

The *long-run variance* of the cost under maintenance policy π is defined as:

$$V(\pi) \equiv \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} [(C_t(\pi) - \Phi(\pi))^2], \tag{1}$$

where

$$\Phi(\pi) \equiv \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} C_t(\pi),$$

is the long-run average cost per unit time under policy π. The variability-sensitive optimization problem is formulated as:

$$\min_{\pi\in K}[(\Phi(\pi))^2 + \lambda V(\pi)], \quad \lambda \in [0, \infty), \tag{2}$$

where K is the class of the considered maintenance polices, and λ is the cost-variability-sensitive factor. $(\Phi(\pi))^2$ is used instead of $\Phi(\pi)$ in Equation (2) to make the units of the two terms consistent. Since $\Phi(\pi) \geq 0$, the policies minimizing $(\Phi(\pi))^2$ must be the same as those minimizing $\Phi(\pi)$. Therefore, Equation (2) is equivalent to the traditional cost-variability-neutral policy when $\lambda = 0$.

In Equation (2), λ can be interpreted as the relative weight of variability as a function of the mean affecting the decision. The intuitive meaning of $\lambda \leq 1$ is that the decision maker considers that the improvement in variability has less impact than the same degree of improvement in the mean cost. In other words, the cost-variability is less significant than the mean cost. Similarly, $\lambda \geq 1$ means that the variability has a more significant impact than the mean cost on the decision-making. We note that the most common intuition of the decision makers is to consider the mean cost to be more significant than the cost-variability. Therefore, the decision makers should be cautious when they decide to use a λ with a value greater than one. Only if the cost-variability is of more concern than the average cost, is a choice of λ with value greater than one reasonable.

Many preventive maintenance problems can be formulated as a renewal process. For renewal processes, the following result on the *long-run variance* of the cost is interesting to note.

Lemma 1. *Let*

$$\Psi(\pi) \equiv \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \left[C_t^2(\pi)\right],$$

which is the long-run mean squared cost. For a renewal process under a certain maintenance policy π, with $C_t(\pi)$ bounded for all $t > 0$:

$$V(\pi) = \Psi(\pi) - (\Phi(\pi))^2,$$

that is, the long-run variance *of the cost is equal to the long-run squared cost per unit time minus the square of the long-run average cost per unit time.*

Proof.

$$\begin{aligned} V(\pi) &= \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} [(C_t(\pi) - \Phi(\pi))^2] \\ &= \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} C_t^2(\pi) \\ &\quad - \lim_{T\to\infty} \frac{2}{T} \sum_{t=1}^{T} C_t(\pi)\Phi(\pi) + \Phi^2(\pi) \\ &= \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} C_t^2(\pi) - \Phi^2(\pi) = \Psi(\pi) - (\Phi(\pi))^2. \end{aligned}$$

For a renewal process with $C_t(\pi)$ bounded, both

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} C_t^2(\pi),$$

and

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} C_t(\pi),$$

above are finite and well defined based on the renewal theory. ■

3. Optimal cost-variability-sensitive maintenance policy for age replacement

3.1. *Long-run average cost in age replacement*

With an age replacement policy, the item is replaced either at failure, or at a specific age a, whichever occurs first. We assume that the cost of a failure/unexpected replacement is c_1 and the cost of a preventive/scheduled replacement is c_2. At each time unit, the item will either continue to work with no cost or be replaced with a new item with a certain cost. Thus, for age replacement:

$$C_t = \begin{cases} c_1, & \text{unexpected/failure replacement at time } t, \\ c_2, & \text{scheduled/preventive replacement at time } t, \\ 0, & \text{the item is not replaced at time } t. \end{cases} \tag{3}$$

The replacement costs should satisfy that $0 < c_2 < c_1$. One of the most common objectives for a variability-neutral decision maker is to minimize the long-run average cost per unit time. Let F be the cumulative distribution function (cdf) of the item's failure time. Under the age replacement policy, the long-run average cost per unit time is (Barlow and Proschan, 1965):

$$\Phi(a) = \frac{c_1 F(a) + c_2 \bar{F}(a)}{S(a)}, \tag{4}$$

where $\Phi(a)$ is the long-run average cost of age replacement with replacement interval a and $S(a) = \int_0^a x\mathrm{d}F(x) + a\int_a^\infty \mathrm{d}F(x)$.

3.2. *Long-run variance of the cost and the variability-sensitive age replacement policy*

The following two assumptions are first made for the variability-sensitive age replacement policies:

Assumption 1. The probability that more than one failure event occurs during one time unit can be ignored.

Assumption 2. $r(t)$ is continuous and increasing (may not be strictly increasing).

Corresponding to Equations (1) and (2), the variability-sensitive criterion for the age replacement policy is:

$$\min_{a\in[1,\infty)} [(\Phi(a))^2 + \lambda V(a)], \quad \lambda \in [0,\infty), \tag{5}$$

where the *long-run variance* of the cost for the age replacement policy with replacement interval a, denoted by $V(a)$, is:

$$V(a) \equiv \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} [(C_t - \Phi(a))^2], \tag{6}$$

where C_t is defined as in Equation (3).

Considering each replacement as a renewal and C_t^2 as the "cost" at each time unit, $\Psi(a)$ can be evaluated based on the renewal theory (Barlow and Proschan, 1965) as:

$$\begin{aligned}\Psi(a) &\equiv \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}C_t^2\\ &= \frac{E[\text{“}cost\text{”}(C_t^2)\ during\ a\ replacement\ cycle]}{E[length\ of\ a\ replacement\ cycle]}\\ &= \frac{c_1^2F(a)+c_2^2\bar{F}(a)}{S(a)}.\end{aligned}$$

Further from Lemma 1, the following result 1 can be easily obtained.

Result 1. For the cost of age replacement defined in Equation (3), the *long-run variance* of the cost for age replacement can be represented by:

$$V(a) = \Psi(a) - \Phi^2(a). \tag{7}$$

Remark 1. From result 1, it can be seen that the definition of the *long-run variance* of the cost in age replacement is analogous to the definition of the variance of a single random variable, that is, $\mathrm{Var}(X) = E(X^2) - E^2(X)$ for a random variable X.

Unlike the variability-neutral case, in general, the solution of the equation:

$$\frac{\mathrm{d}}{\mathrm{d}a}((\Phi(a))^2 + \lambda V(a)) = 0,$$

is not unique and the optimal solution to the optimization problem stated in Equation (5) is not necessarily unique even if the failure rate $r(t)$ is continuous and strictly increasing. To find the optimal solution(s), a search through $a = 1, 2, 3, \ldots$ is needed to find the a^* to minimize $(\Phi(a))^2 + \lambda V(a)$. The following result 2 on the boundary of the optimal solution is used to facilitate this search.

Lemma 2. *For* $a > 0$:

$$\frac{\mathrm{d}\Psi(a)}{\mathrm{d}a} = (c_1+c_2)\frac{\mathrm{d}\Phi(a)}{\mathrm{d}a} + \frac{c_1c_2\bar{F}(a)}{S^2(a)}. \tag{8}$$

Proof.

$$\frac{\mathrm{d}\Psi(a)}{\mathrm{d}a} = \frac{r(a)(c_1^2-c_2^2)S(a) - F(a)(c_1^2-c_2^2) - c_2^2}{S^2(a)/\bar{F}(a)}, \tag{9}$$

$$\frac{\mathrm{d}\Phi(a)}{\mathrm{d}a} = \frac{r(a)(c_1-c_2)S(a) - F(a)(c_1-c_2) - c_2}{S^2(a)/\bar{F}(a)}. \tag{10}$$

Equation (8) is followed by Equations (9) and (10). ■

Let $A(\Phi)$ denote the set of all optimal solutions which minimize $\Phi(a)$. $A(\lambda_1)$ and $A(\lambda_2)$ are used to denote the sets of all optimal solutions which minimize the optimization problem in Equation (5) with $\lambda = \lambda_1$ and $\lambda = \lambda_2$, respectively. Note that infinity can be an element in all the above defined sets.

Result 2. Let $a_\Phi^* \equiv \inf A(\Phi)$, and a^* be an optimal solution of the optimization problem in Equation (5) with a fixed λ. If $\lambda > 0$, then $a^* \le a_\Phi^*$. The inequality is strict if $a_\Phi^* < \infty$. Furthermore, let $\lambda_1 > \lambda_2 \ge 0$, $a_{\lambda_1}^* \equiv \inf A(\lambda_1)$, and $a_{\lambda_2}^* \equiv \inf A(\lambda_2)$, then $a_{\lambda_1}^* \le a_{\lambda_2}^*$.

The detailed proof of result 2 is given in Appendix 1.

From result 2, the following two conclusions can be obtained:

- Result 2 provides a search bound for the optimal solution, that is, the search will only be conducted within the interval $[1, a_\Phi^*]$, which significantly reduces the search region. The traditional optimal replacement policy a_Φ^* is generally easier to obtain than a^* because it is known that under assumption 2 the solution of $\mathrm{d}\Phi(a)/\mathrm{d}a = 0$ must be the global optimum for $\Phi(a)$ (Barlow and Proschan, 1965).
- The optimal replacement interval under the variability-sensitive policy tends to be shorter than that under the traditional variability-neutral policy. In other words, the variability-sensitive optimal policy tends to be more conservative than the traditional optimal policy. In addition, the greater the cost-variability-sensitivity (corresponding to a larger λ), the shorter is the preventive replacement interval. This seems to be consistent with the intuition that the maintenance risk can be reduced through increasing the replacement frequency, i.e., the more frequently the items are replaced, the lower will be the cost variability. However, the strict proof of result 2 is not trivial, and it is presented in Appendix 1. A numerical example that will be presented in Section 5, will show that, contrary to intuition, the cost variability is not simply a monotone increasing function over replacement intervals.

From Barlow and Proschan (1965), if $r(t)$ is continuous and strictly increasing to infinity, $a_\Phi^* < \infty$. Thus, the following corollary of result 2 is obviously followed.

Corollary 1. *There exists a finite optimal solution for the optimization problem in Equation (5) if $r(t)$ is continuous and strictly increasing to infinity.*

4. Two other preventive maintenance policies

In addition to the age replacement policy discussed above, there are some other popular preventive maintenance policies in the literature such as periodic replacement under minimal repair (Barlow and Proschan, 1965) and replacement policy N (Lam, 1990). The variability-sensitive maintenance policies for both periodic replacement under minimal repair and replacement policy N will be discussed in this section.

4.1. *Periodic replacement under minimal repair*

In the maintenance policy of periodic replacement under minimal repair a replacement is performed at the fixed times $a, 2a, 3a, \ldots (a > 0)$ with a replace cost c_r; while the minimal repair is performed with a cost c_m for any intervening failure. The replacement cycle under this policy has an equal length of a. The maintenance decision is to select the optimal replacement interval a.

Similar to the analysis for age replacement, the *long-run variance* of the cost for periodic replacement under minimal repair, denoted by $V_I(a)$, is defined as:

$$V_I(a) \equiv \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T}[(C_t - \Phi_I(a))^2], \tag{11}$$

where $\Phi_I(a)$ is the long-run average cost of periodic replacement under minimal repair. The optimization problem is defined as:

$$\min_{a\in(1,\infty)} [(\Phi_I(a))^2 + \lambda V_I(a)], \quad \lambda \in [0, \infty). \tag{12}$$

In addition to assumptions 1 and 2, the following assumption is made for periodic replacement under minimal repair.

Assumption 3. For any meaningful policy a, we assume that:

$$\Phi_I(a) \le \frac{\max(c_r, c_m)}{2}.$$

This condition is generally true except for the extreme maintenance decisions under which a replacement or a failure occurs every two time units. The reason for these exceptions is that the defined length of the basic time units is too large. Thus, the time unit should be reduced to a shorter length.

Since periodic replacement under minimal repair can also be considered as a renewal process, Lemma 1 also holds in this case. From the renewal theory, the long-run average cost for periodic replacement under minimal repair can be obtained as (Barlow and Proschan, 1965):

$$\Phi_I(a) = \frac{c_m \int_0^a r(u)\mathrm{d}u + c_r}{a}.$$

And the *long-run variance* of the cost can be stated by result 3.

Result 3. The *long-run variance* of the cost for periodic replacement under minimal repair can be written as

$$V_I(a) = \Psi_I(a) - [\Phi_I(a)]^2, \tag{13}$$

where

$$\Psi_I(a) \equiv \frac{c_m^2 \int_0^a r(u)\mathrm{d}u + c_r^2}{a}.$$

Let $B(\Phi_I)$ represent the set of optimal solutions which minimize $\Phi_I(a)$, and $B(\lambda_1)$ and $B(\lambda_2)$ denote the sets of optimal solutions to the optimization problem in Equation (12) with $\lambda = \lambda_1$ and $\lambda = \lambda_2$, respectively. The bound of the optimal variability-sensitive periodic replacement policy under minimal repair and the effects of the variability sensitivity on the optimal policy can be seen from result 4.

Result 4. Let a_I^* be an optimal solution of Equation (12) with $\lambda \ge 0$, $b_{\inf}^* \equiv \inf B(\Phi_I)$, $b_{\sup}^* \equiv \sup B(\Phi_I)$; $\lambda_1 > \lambda_2 \ge 0$, $a_{I,\lambda_1}^* \equiv \inf B(\lambda_1)$, and $a_{I,\lambda_2}^* \equiv \inf B(\lambda_2)$. If $c_m > c_r$ (or $c_m < c_r$), then $a_I^* \le b_{\inf}^*$ (or $a_I^* \ge b_{\sup}^*$) and $a_{I,\lambda_1}^* \le a_{I,\lambda_2}^*$(or $a_{I,\lambda_1}^* \ge a_{I,\lambda_2}^*$).

The detailed proof of result 4 can be found in Appendix 2.

Remark 2. Result 4 indicates that the variability-sensitive optimal policy tends to be more conservative than the traditional variability-neutral optimal policy.

(i) When the minimal repair cost is higher than the replacement cost, the variability-sensitive policy tends to increase the replacement frequency. And the greater the cost-variability-sensitivity (corresponding to a larger λ), the higher the preventive replacement frequency.
(ii) Contrary to (i), when the replacement cost is higher than the minimal repair cost, the variability-sensitive policy decreases the replacement frequency, leading to more chances for minimal repairs. Also the greater the cost-variability-sensitivity, the lower the preventive replacement frequency.

Corollary 2. *If $c_m < c_r$, the necessary condition for the variability-sensitive periodic replacement policy to have a finite optimal solution is that the variability-neutral problem has a finite optimal solution. If $c_m > c_r$, the sufficient condition for the variability-sensitive periodic replacement policy to have a finite optimal solution is that the variability-neutral problem has a finite optimal solution.*

4.2. *Replacement policy N*

The replacement policy N performs minimal repairs for the first $N - 1$ failures and replaces the Nth failure. This policy has been considered by Park (1979) and Nakagawa (1984). In order to study variability-sensitive maintenance under this maintenance policy, a more general replacement model based on the following assumptions is developed by extending Lam (1990):

Assumption 4. When an item fails, it is either repaired or replaced. The decision to replace will be determined according to replacement policy N.

Assumption 5. Let X_k be the failure time after the $(k-1)$th repair, then $\{X_k, k = 1, 2, \ldots\}$ forms a sequence of non-negative random variables with non-increasing means $E(X_k) = \mu_k$.

Assumption 6. Let Y_k be the repair time (time spent on a repair) after the kth failure, then $\{Y_k, k = 1, 2, \ldots\}$ constitutes a sequence of non-negative random variables with non-decreasing means $E(Y_k) = \eta_k$.

Assumption 7. We extend Lam (1990) by considering the replacement time (time spent on a replacement). Let random variable Z be the replacement time whose mean is $E(Z) = \gamma > 0$.

Assumption 8. The values for X_k, Y_k, and Z are all integers.

Assumption 9. The repair cost per unit time is c_m; the replacement cost per unit time under replacement policy N is c_r. Both c_m and c_r include the loss of earnings/rewards for operating the system.

Assumption 10. For all concerned policies, the availability of the system is greater than 50%. This assumption is generally true for most real applications. If the obtained optimal policy gives a system availability of less than 50%, then either the system is inadequate and a system redesign is required or the optimization criterion is incorrect. A sufficient condition for this assumption is $\mu_k \geq \max[\eta_k, \gamma], \forall k$.

Similar to the age replacement policy, the *long-run variance* of the cost for replacement policy N, denoted by $V_{\mathrm{II}}(N)$, can be defined as:

$$V_{\mathrm{II}}(N) \equiv \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T}[(C_t - \Phi_{\mathrm{II}}(N))^2], \quad (14)$$

where $\Phi_{\mathrm{II}}(N)$ is the long-run average cost of replacement policy N. The optimization problem is defined as:

$$\min_{N\in[1,\infty)} [(\Phi_{\mathrm{II}}(N))^2 + \lambda V_{\mathrm{II}}(N)], \quad \lambda \in [0,\infty). \quad (15)$$

Since replacement policy N results in a renewal process, Lemma 1 also holds for replacement policy N. From the renewal theory, the long-run average cost under replacement policy N can be obtained as:

$$\Phi_{\mathrm{II}}(N) = \frac{c_m \sum_{k=1}^{N-1}\eta_k + c_r\gamma}{\sum_{k=1}^{N}\mu_k + \sum_{k=1}^{N-1}\eta_k + \gamma}.$$

And the *long-run variance* of the cost can be stated by result 5.

Result 5. The *long-run variance* of the cost for replacement policy N can be written as:

$$V_{\mathrm{II}}(N) = \Psi_{\mathrm{II}}(N) - (\Phi_{\mathrm{II}}(N))^2, \quad (16)$$

where

$$\Psi_{\mathrm{II}}(N) \equiv \frac{c_m^2 \sum_{k=1}^{N-1}\eta_k + c_r^2\gamma}{\sum_{k=1}^{N}\mu_k + \sum_{k=1}^{N-1}\eta_k + \gamma}.$$

Similar to the arguments in Lam (1990), it is easy to see that:

$$\Phi_{\mathrm{II}}(N+1) - \Phi_{\mathrm{II}}(N) = (c_m f_N - c_r\gamma(\mu_{N+1} + \eta_N))/\Delta_N,$$

where

$$\Delta_N = \left(\sum_{k=1}^{N+1}\mu_k + \sum_{k=1}^{N}\eta_k + \gamma\right)\left(\sum_{k=1}^{N}\mu_k + \sum_{k=1}^{N-1}\eta_k + \gamma\right),$$

and

$$f_N = \eta_N \sum_{k=1}^{N}\mu_k - \mu_{N+1}\sum_{k=1}^{N-1}\eta_k.$$

Define $g_N \equiv \gamma(\mu_{N+1} + \eta_N)/f_N$, we have that:

1. $\{g_N\}$ is a non-increasing sequence;
2. $\Phi_{\mathrm{II}}(N+1) \leq (\geq)\Phi_{\mathrm{II}}(N)$ iff $g_N \geq (\leq) c_m/c_r$;
3. the variability-neutral optimal policy N_0^* is determined by:

$$N_0^* = \min\{N \geq 1 \mid g_N \leq c_m/c_r\}.$$

For the variability-sensitive optimal policy, let $C(\Phi_{\mathrm{II}})$, $C(\lambda_1)$, and $C(\lambda_2)$ be the sets of optimal solutions minimizing $\Phi_{\mathrm{II}}(N)$, the optimization problem in Equation (15) with $\lambda = \lambda_1$, and the optimization problem in Equation (15) with $\lambda = \lambda_2$, respectively. The bound of the optimal variability-sensitive replacement policy N and the effect of the variability sensitivity on the optimal policy can be seen from result 6.

Result 6. Let N^* be the optimal variability-sensitive replacement policy with a fixed $\lambda \geq 0$, $c_{\inf}^* \equiv \inf C(\Phi_{\mathrm{II}})$, $c_{\sup}^* \equiv \sup C(\Phi_{\mathrm{II}})$, $N_{\lambda_1}^* \equiv \inf C(\lambda_1)$, and $N_{\lambda_2}^* \equiv \inf C(\lambda_2)$. If $\lambda_1 > \lambda_2 \geq 0$ and $c_m > c_r$ (or $c_m < c_r$), then $N^* \leq c_{\inf}^*$ (or $N^* \geq c_{\sup}^*$) and $N_{\lambda_1}^* \leq N_{\lambda_2}^*$ (or $N_{\lambda_1}^* \geq N_{\lambda_2}^*$).

The detailed proof of result 6 can be found in Appendix 3.

Similarly, result 6 suggests that the variability-sensitive optimal policy is more conservative than the traditional policy.

Corollary 3. *If $c_m < c_r$, the necessary condition for the variability-sensitive replacement policy N to have a finite optimal solution is that the variability-neutral problem has a finite optimal solution. If $c_m > c_r$, the sufficient condition for the variability-sensitive replacement policy N to have a finite optimal solution is that the variability-neutral problem has a finite optimal solution.*

5. Numerical example

The age replacement policy is one of the most popular preventive maintenance policies currently used in industry. An example of a spot welding gun system is given to illustrate the variability-sensitive age replacement policy. It is assumed that the transition of the machine condition is subject to the Markov chain model illustrated in Fig. 1 where state 1 is the good operating state, state 2 is the deteriorated state, and state 3 is the failure state. At the deteriorated state 2, a cooling subsystem failure, such as leakage, leads to significant performance deterioration of the welding gun system, although the spot welding gun may still work in this state. Therefore, the failure rate of the welding gun will be significantly increased at state 2. The sojourn

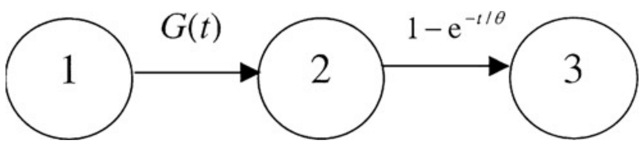


Fig. 1. A Markov chain failure model.

time distribution at state 1 follows a lognormal distribution with cdf $G(t)$. The sojourn time at state 2 is exponential distributed with mean θ. State 3 is an absorbing state. The initial failure rate of a welding gun under the good condition is much smaller than that when the cooling system fails. Thus, the initial failure rate is ignored in the following analysis and the failure process of a spot welding gun can be modeled with the Markov chain as shown in Fig. 1. Table 1 gives the parameters used in the Markov chain model.

The age replacement policy is used for the maintenance of this spot welding gun. The analysis results are given in Table 2 to compare the variability-sensitive optimal replacement policies with the traditional variability-neutral replacement policy. The decision is based on the integer values of the replacement interval.

From Table 2, it can be seen that the optimal replacement intervals under the variability-sensitive optimal policies are quite different from that under the traditional variability-neutral optimal policy. The traditional policy decides that the replacement should be performed only at failure time ($a^* = +\infty$), while the variability-sensitive replacement policy suggests a preventive replacement at an early age ($a^* = 5$ or 6 weeks). Comparing the long-run average costs, all three policies have close values of $\Phi(a^*)$ as shown in Table 2. The variability-sensitive policy results in only 5% (or 0.5%) higher values of the long-run average cost than the traditional policy for $\lambda = 0.2$ (or $\lambda = 0.02$). However, the *long-run variance* of the cost under the variability-neutral policy is three to four times greater than that under the variability-sensitive policies. Therefore, this example shows that the variability-sensitive maintenance policy can significantly reduce the cost variation with only a very small increase of the long-run average cost.

Figure 2 is used to further understand the optimality of the variability-sensitive maintenance policies. Policies A, B and C designated in Fig. 2 are used to represent three typical maintenance policies with the replacement intervals being less than 5 weeks, equal to 5 or 6 weeks, and greater than 6 weeks, respectively. When the preventive replacement interval is greater than 5 weeks (policies B and C), the long-run average cost ($\Phi(a)$) is decreasing but quite flat over the replacement interval a ($a > 5$). So from the average cost point of view, there is no significant difference to choose between policies B and C. However, the cost variability is increased significantly from policy B to policy C. Therefore, considering both the cost variability and the average cost, policy B should be chosen as the optimal policy so that the cost variability is much smaller than policy C and the average cost is still close to the minimum. Figure 2 also shows that the cost variability is not a monotone increasing function over the replacement interval a, that is, decreasing the replacement interval does not necessarily always reduces the cost variability. However, it is proved in previous sections that the variability-sensitive replacement policies are more conservative than the traditional optimal

Table 1. Parameters in the Markov chain model

Parameter name	*Value*
Time unit	1 week
Mean sojourn time at state 1 (week)	5
Standard deviation of sojourn time at state 1 (week)	0.5
Mean sojourn time at state 2 (week)	25
Scheduled replacement cost ($000)	1
Unscheduled replacement cost ($000)	6

Table 2. Comparison between variability-sensitive policy and variability-neutral policy

	Optimal replacement time (a^*)	$\Phi(a^*)$	$V(a^*)$	$(\Phi(a^*))^2 + \lambda V(a^*)$
Variability-neutral policy	$+\infty$	0.200	1.160	0.272 ($\lambda = 0.2$) 0.063 ($\lambda = 0.02$)
Variability-sensitive policy ($\lambda = 0.2$)	5	0.209	0.219	0.088
Variability-sensitive policy ($\lambda = 0.02$)	6	0.2009	0.362	0.048

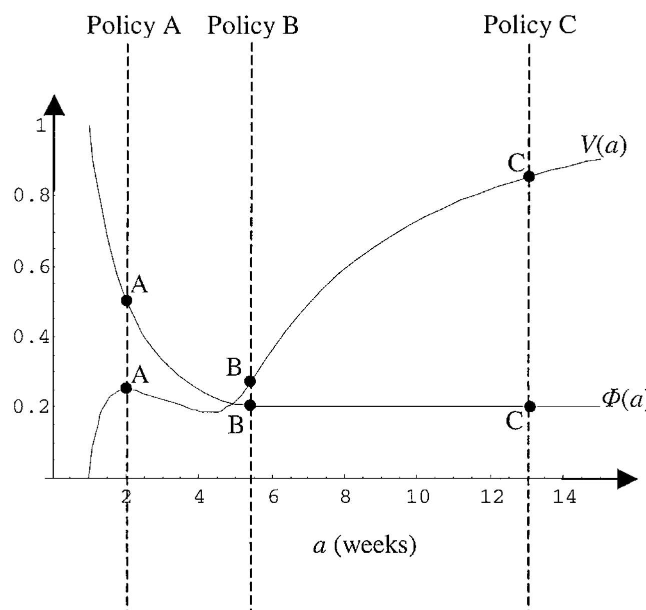
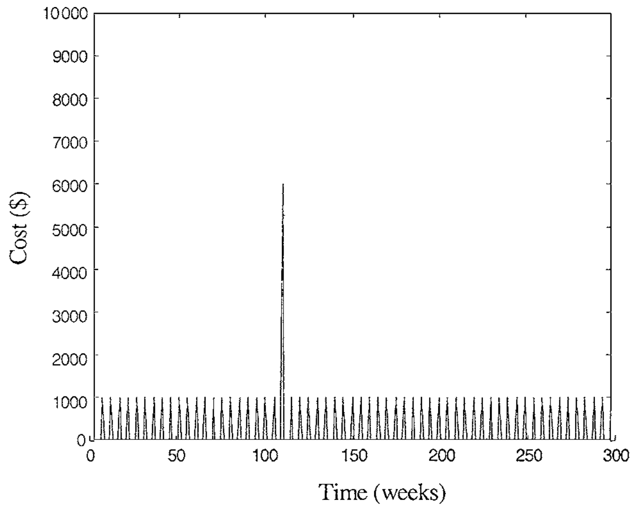
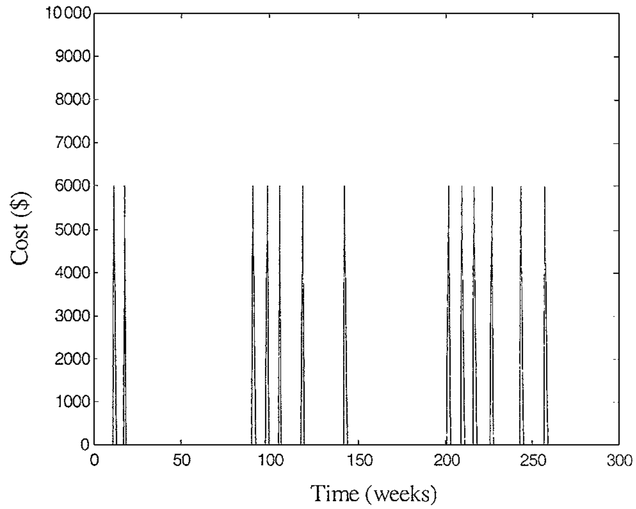


Fig. 2. Average cost and cost variation for different replacement intervals.



(a)



(b)

Fig. 3. Simulated cost streams under different policies: (a) the cost streams under the variability-sensitive policy; and (b) the cost streams under the variability-neutral policy.

policy, that is, variability-sensitive optimal policies will always have smaller replacement intervals than the respective variability-neutral optimal policies.

Figure 3(a and b) compares the simulated cost streams under the variability-sensitive policy and the variability-neutral policy. This cost stream is one of the 10 000 generated cost streams using MATLAB. From Fig. 3(a and b) it is easy to see that the variability-sensitive policy has a more stable nature for the cost stream, while the variability-neutral policy possesses a more unpredictable nature for the cost stream. Furthermore, under the variability-neutral policy, highly unexpected costs can be found during the time periods from 100 to 150 weeks and from 200 to 250 weeks, which will lead to severe potential risks for maintenance management under a given budget allocation. Among the 10 000 cost streams generated, the variability-sensitive policy leads to 0.54 unexpected/failure replacements per 300 weeks of operations while variability-neutral policy leads to 9.87 unexpected/failure replacements per 300 weeks of operations. Therefore, Fig. 3(a and b), showing one unexpected/failure replacement for the variability-sensitive policy and 13 unexpected/failure replacements for the variability-neutral policy, can be considered as a typical pattern of the cost streams for this example.

6. Conclusions and future work

A general analysis framework of the cost-variability-sensitive maintenance decision-making is proposed in this paper. Based on this framework, three popular preventive maintenance policies: (i) age replacement; (ii) periodic replacement under minimal repair; and (iii) replacement policy N are reinvestigated. The bounds of the optimal solutions for each maintenance policy have been proved. To avoid the management risk created by maintenance activities with unexpected higher costs, it is shown that the optimal variability-sensitive policy always tends to conduct maintenance more frequently with a lower cost for each action to avoid the risk of maintenance actions with a higher cost. In this sense, the variability-sensitive policies are more conservative than the traditional maintenance policies. This interesting property also explicates the search bound for the optimal policy. Furthermore, the greater the cost-variability-sensitivity (corresponding to a larger λ in the optimization formulations), the more conservative will be the optimal variability sensitive policy. An example for the age replacement policy is given to illustrate the importance of considering the cost variability effect on risk management. Compared with the traditional optimal maintenance policy, the numerical solution shows that the proposed variability-sensitive optimal policies can significantly reduce the maintenance management risk with only a little increase in the expected cost.

The new variability-sensitive maintenance framework developed in this paper for some typical preventive maintenance policies is general in nature and the potential benefits of the variability-sensitive maintenance policy as compared to traditional maintenance policies can be seen clearly from the numerical example. In this paper, the cost components are assumed to be known constants. However, in some applications there is significant uncertainty in maintenance costs. Therefore, an interesting future development would be to capture the cost uncertainty in the proposed framework. Also of interest is the extension of the variability-sensitive maintenance framework to more complex maintenance polices such as condition-based preventive maintenance polices where more factors such as inspection cost and the failure detection capability can be considered in the more complex structure of the cost streams.

Acknowledgements

The authors thank the associate editor and the referees for their valuable comments and suggestions. In this research, Dr. Y. Chen and Dr. J. Jin are partially supported by the National Science Foundation (NSF) under grant NSF-DMI-9713654, and CAREER AWARD NSF DMI-0133942.

References

Barlow, R.E. and Proschan, F. (1965) *Mathematical Theory of Reliability*, John Wiley, New York.

Filar, J.A., Kallenberg, L.C.M., and Lee, H.-M. (1989) Variance-penalized Markov decision processes. *Mathematics of Operations Research*, **14**(1), 147–161.

Filar, J.A. and Lee, H.M. (1984) Gain/variability tradeoffs in undiscounted Markov decision processes, in *Proceedings of the 1985 IEEE Conference on Decision and Control*, IEEE, Piscataway, NJ, pp. 1106–1112.

Fox, B. (1966) Age replacement with discounting. *Operations Research*, **14**, 533–537.

Glasser, G.J. (1967) The age replacement problem. *Technometrics*, **9**(1), 83–91.

Lam, Y. (1990) A repair replacement model. *Advances in Applied Probability*, **22**(2), 494–497.

Liu, P.H., Makis, V., and Jardine, A.K.S. (2001) Scheduling of the optimal tool replacement times in a flexible manufacturing system. *IIE Transactions*, **33**(6), 487–495.

Markowitz, H. (1959) *Portfolio Selection*, Wiley, New York, NY.

Nakagawa, T. (1984) *Discrete Replacement Models*, Lecture Notes in Economics and Mathematical Systems, 235, Springer-Verlag, Berlin, Germany.

Nakagawa, T. and Osaki, S. (1974) The optimal repair limit replacement policies. *Operations Research*, **25**, 311–317.

Park, K.S. (1979) Optimal number of minimal repairs before replacement. *IEEE Transactions On Reliability*, **28**, 137–140.

Rangan, A. and Grace, R.E. (1988) A non-Markov model for the optimum replacement of self-repairing systems subject to shocks. *Journal of Applied Probability*, **25**, 375–382.

Sobel, M.J. (1985) Maximal mean/standard deviation ratio in an undiscounted MDP. *Operations Research Letters*, **4**, 157–159.

Sobel, M.J. (1994) Mean-variance tradeoffs in an undiscounted MDP. *Operations Research*, **42**(1), 175–183.

Tapiero, C.S. and Venezia, I. (1979) A mean variance approach to the optimal machine maintenance and replacement. *OR; The Journal of the Operational Research Society*, **30**(5), 457–466.

White, D.J. (1988) Mean, variance, and probabilistic criteria in finite Markov decision processes: a review. *Journal of Optimization Theory and Applications*, **56**(1), 1–29.

Zhang, F. and Jardine, A.K.S. (1998) Optimal maintenance models with minimal repair, periodic overhaul and complete renewal. *IIE Transactions*, **30**(12), 1109–1119.

Appendices

Appendix 1: Proof of result 2

To prove $a^* < a^*_\Phi$ when $a^*_\Phi < \infty$, first note that from result 1:

$$\frac{\mathrm{d}}{\mathrm{d}a}((\Phi(a))^2 + \lambda V(a)) = (2\Phi(a) - 2\lambda\Phi(a))\frac{\mathrm{d}\Phi(a)}{\mathrm{d}a} + \lambda\frac{\mathrm{d}\Psi(a)}{\mathrm{d}a}. \quad \text{(A1)}$$

From Lemma 2 and Equation (A1):

$$\frac{\mathrm{d}}{\mathrm{d}a}((\Phi(a))^2 + \lambda V(a)) = (2\Phi(a) + \lambda(c_1 + c_2) - 2\lambda\Phi(a))\frac{\mathrm{d}\Phi(a)}{\mathrm{d}a} + \frac{\lambda c_1 c_2 \bar{F}(a)}{S^2(a)}. \quad \text{(A2)}$$

It is easy to show that for $a^*_\Phi < \infty$:

$$\forall a \geq a^*_\Phi, \quad \frac{\mathrm{d}\Phi(a)}{\mathrm{d}a} \geq 0. \quad \text{(A3)}$$

Also since $\mu \geq 2$ and $\forall a \geq a^*_\Phi$, we have:

$$\Phi(a) \leq \Phi(\infty) = \frac{c_1}{\mu} < \frac{c_1 + c_2}{2}. \quad \text{(A4)}$$

Equations (A2), (A3) and (A4) result that if $\lambda > 0$:

$$\frac{\mathrm{d}}{\mathrm{d}a}((\Phi(a))^2 + \lambda V(a)) > 0, \quad \forall a \geq a^*_\Phi.$$

Hence, $a^* < a^*_\Phi$ when $a^*_\Phi < \infty$.

To prove $a^*_{\lambda_1} \leq a^*_{\lambda_2}$ with $\lambda_1 > \lambda_2 \geq 0$, notice that $a^*_{\lambda_1} \leq a^*_\Phi$ and $a^*_{\lambda_2} \leq a^*_\Phi$. It is easy to see that

$$\forall a \leq a^*_\Phi, \quad \frac{\mathrm{d}\Phi(a)}{\mathrm{d}a} \leq 0. \quad \text{(A5)}$$

Contrary to the conclusion, we suppose that $a^*_{\lambda_1} > a^*_{\lambda_2}$. Then, $\Phi(a^*_{\lambda_1}) < \Phi(a^*_{\lambda_2})$. Further from:

$$\left(\Phi\left(a^*_{\lambda_2}\right)\right)^2 + \lambda_2 V\left(a^*_{\lambda_2}\right) \leq \left(\Phi\left(a^*_{\lambda_1}\right)\right)^2 + \lambda_2 V\left(a^*_{\lambda_1}\right),$$

we have that $V(a^*_{\lambda_2}) < V(a^*_{\lambda_1})$. Thus:

$$\begin{aligned} 0 &\leq \left(\Phi\left(a^*_{\lambda_1}\right)\right)^2 + \lambda_2 V\left(a^*_{\lambda_1}\right) - \left(\left(\Phi\left(a^*_{\lambda_2}\right)\right)^2 + \lambda_2 V\left(a^*_{\lambda_2}\right)\right) \\ &= \left(\Phi\left(a^*_{\lambda_1}\right)\right)^2 - \left(\Phi\left(a^*_{\lambda_2}\right)\right)^2 + \lambda_2\left(V\left(a^*_{\lambda_1}\right) - V\left(a^*_{\lambda_2}\right)\right) \\ &< \left(\Phi\left(a^*_{\lambda_1}\right)\right)^2 - \left(\Phi\left(a^*_{\lambda_2}\right)\right)^2 + \lambda_1\left(V\left(a^*_{\lambda_1}\right) - V\left(a^*_{\lambda_2}\right)\right). \end{aligned}$$

That is, $(\Phi(a^*_{\lambda_1}))^2 + \lambda_1 V(a^*_{\lambda_1}) > (\Phi(a^*_{\lambda_2}))^2 + \lambda_1 V(a^*_{\lambda_2})$, which contradicts with the fact that $(\Phi(a))^2 + \lambda_1 V(a)$ achieves a minimum at $a^*_{\lambda_1}$. Therefore, it can be concluded that $a^*_{\lambda_1} \leq a^*_{\lambda_2}$. ■

Appendix 2: Proof of result 4

When $c_\mathrm{m} > c_\mathrm{r}$:

$$\begin{aligned} \frac{\mathrm{d}\Psi_\mathrm{I}(a)}{\mathrm{d}a} &= \frac{c_\mathrm{m}^2 \int_0^a [r(a) - r(u)]\mathrm{d}u - c_\mathrm{r}^2}{a^2} \\ &> \frac{c_\mathrm{m}\left(c_\mathrm{m} \int_0^a [r(a) - r(u)]\mathrm{d}u - c_\mathrm{r}\right)}{a^2} \\ &= c_\mathrm{m}\frac{\mathrm{d}\Phi_\mathrm{I}(a)}{\mathrm{d}a}. \end{aligned}$$

From result 3:

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}a}((\Phi_\mathrm{I}(a))^2 + \lambda V_\mathrm{I}(a)) \\ &= (2\Phi_\mathrm{I}(a) - 2\lambda\Phi_\mathrm{I}(a))\frac{\mathrm{d}\Phi_\mathrm{I}(a)}{\mathrm{d}a} + \lambda\frac{\mathrm{d}\Psi_\mathrm{I}(a)}{\mathrm{d}a} \\ &> (2\Phi_\mathrm{I}(a) + \lambda c_\mathrm{m} - 2\lambda\Phi_\mathrm{I}(a))\frac{\mathrm{d}\Phi_\mathrm{I}(a)}{\mathrm{d}a}. \end{aligned}$$

By assumption 3, $\Phi_\mathrm{I}(a) \le c_\mathrm{m}/2$. Also, it is easy to see that if $b^*_{\inf} < \infty$, $\forall a \ge b^*_{\inf}$, we have $\mathrm{d}\Phi_\mathrm{I}(a)/\mathrm{d}a \ge 0$. Hence, $\forall a \ge b^*_{\inf}$ we also have:

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}a}((\Phi_\mathrm{I}(a))^2 + \lambda V_\mathrm{I}(a))\\ &\quad > (2\Phi_\mathrm{I}(a) + \lambda c_\mathrm{m} - 2\lambda\Phi_\mathrm{I}(a))\frac{\mathrm{d}\Phi_\mathrm{I}(a)}{\mathrm{d}a} \ge 0,\end{aligned}$$

from which it follows that $a^*_\mathrm{I} < b^*_{\inf}$.

To prove $a^*_{\mathrm{I},\lambda_1} \le a^*_{\mathrm{I},\lambda_2}$ with $\lambda_1 > \lambda_2 \ge 0$, note that $a^*_{\mathrm{I},\lambda_1} \le b^*_{\inf}$ and $a^*_{\mathrm{I},\lambda_2} \le b^*_{\inf}$. It is easy to see that:

$$\forall a \le b^*_{\inf},\ \frac{\mathrm{d}\Phi_\mathrm{I}(a)}{\mathrm{d}a} \le 0. \tag{A6}$$

Contrary to the conclusion, we suppose that $a^*_{\mathrm{I},\lambda_1} > a^*_{\mathrm{I},\lambda_2}$. Then, $\Phi_\mathrm{I}(a^*_{\mathrm{I},\lambda_1}) < \Phi_\mathrm{I}(a^*_{\mathrm{I},\lambda_2})$. Further from:

$$\left(\Phi_\mathrm{I}\left(a^*_{\mathrm{I},\lambda_2}\right)\right)^2 + \lambda_2 V_\mathrm{I}\left(a^*_{\mathrm{I},\lambda_2}\right) \le \left(\Phi_\mathrm{I}\left(a^*_{\mathrm{I},\lambda_1}\right)\right)^2 + \lambda_2 V_\mathrm{I}\left(a^*_{\mathrm{I},\lambda_1}\right),$$

we have that $V_\mathrm{I}(a^*_{\mathrm{I},\lambda_2}) < V_\mathrm{I}(a^*_{\mathrm{I},\lambda_1})$. Thus:

$$\begin{aligned}0 &\le \left(\Phi_\mathrm{I}\left(a^*_{\mathrm{I},\lambda_1}\right)\right)^2 + \lambda_2 V_\mathrm{I}\left(a^*_{\mathrm{I},\lambda_1}\right) - \left(\left(\Phi_\mathrm{I}\left(a^*_{\mathrm{I},\lambda_2}\right)\right)^2 + \lambda_2 V_\mathrm{I}\left(a^*_{\mathrm{I},\lambda_2}\right)\right)\\ &= \left(\Phi_\mathrm{I}\left(a^*_{\mathrm{I},\lambda_1}\right)\right)^2 - \left(\Phi_\mathrm{I}\left(a^*_{\mathrm{I},\lambda_2}\right)\right)^2 + \lambda_2\left(V_\mathrm{I}\left(a^*_{\mathrm{I},\lambda_1}\right) - V_\mathrm{I}\left(a^*_{\mathrm{I},\lambda_2}\right)\right)\\ &< \left(\Phi_\mathrm{I}\left(a^*_{\mathrm{I},\lambda_1}\right)\right)^2 - \left(\Phi_\mathrm{I}\left(a^*_{\mathrm{I},\lambda_2}\right)\right)^2 + \lambda_1\left(V_\mathrm{I}\left(a^*_{\mathrm{I},\lambda_1}\right) - V_\mathrm{I}\left(a^*_{\mathrm{I},\lambda_2}\right)\right).\end{aligned}$$

That is, $(\Phi_\mathrm{I}(a^*_{\mathrm{I},\lambda_1}))^2 + \lambda_1 V_\mathrm{I}(a^*_{\mathrm{I},\lambda_1}) > (\Phi_\mathrm{I}(a^*_{\mathrm{I},\lambda_2}))^2 + \lambda_1 V_\mathrm{I}(a^*_{\mathrm{I},\lambda_2})$, which contradicts with the fact that $(\Phi_\mathrm{I}(a))^2 + \lambda_1 V_\mathrm{I}(a)$ achieves a minimum at $a^*_{\mathrm{I},\lambda_1}$. Therefore, it can be concluded that $a^*_{\mathrm{I},\lambda_1} \le a^*_{\mathrm{I},\lambda_2}$.

When $c_\mathrm{m} < c_\mathrm{r}$:

$$\begin{aligned}\frac{\mathrm{d}\Psi_\mathrm{I}(a)}{\mathrm{d}a} &= \frac{c_\mathrm{m}^2\int_0^a [r(a) - r(u)]\mathrm{d}u - c_\mathrm{r}^2}{a^2}\\ &< \frac{c_\mathrm{r}\left(c_\mathrm{m}\int_0^a [r(a) - r(u)]\mathrm{d}u - c_\mathrm{r}\right)}{a^2}\\ &= c_\mathrm{r}\frac{\mathrm{d}\Phi_\mathrm{I}(a)}{\mathrm{d}a}.\end{aligned}$$

From result 3:

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}a}((\Phi_\mathrm{I}(a))^2 + \lambda V_\mathrm{I}(a))\\ &\quad = (2\Phi_\mathrm{I}(a) - 2\lambda\Phi_\mathrm{I}(a))\frac{\mathrm{d}\Phi_\mathrm{I}(a)}{\mathrm{d}a} + \lambda\frac{\mathrm{d}\Psi_\mathrm{I}(a)}{\mathrm{d}a}\\ &\quad < (2\Phi_\mathrm{I}(a) + \lambda c_\mathrm{r} - 2\lambda\Phi_\mathrm{I}(a))\frac{\mathrm{d}\Phi_\mathrm{I}(a)}{\mathrm{d}a}.\end{aligned}$$

By assumption 3, $\Phi_\mathrm{I}(a) \le c_\mathrm{r}/2$. Also, it is easy to see that $\forall a \le b^*_{\sup}$, we have $\mathrm{d}\Phi_\mathrm{I}(a)/\mathrm{d}a \le 0$. Hence, $\forall a \le b^*_{\sup}$ we also have:

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}a}((\Phi_\mathrm{I}(a))^2 + \lambda V_\mathrm{I}(a))\\ &\quad < (2\Phi_\mathrm{I}(a) + \lambda c_\mathrm{m} - 2\lambda\Phi_\mathrm{I}(a))\frac{\mathrm{d}\Phi_\mathrm{I}(a)}{\mathrm{d}a} \le 0,\end{aligned}$$

from which it follows that $a^*_\mathrm{I} > b^*_{\sup}$.

To prove $a^*_{\mathrm{I},\lambda_1} \ge a^*_{\mathrm{I},\lambda_2}$ with $\lambda_1 > \lambda_2 \ge 0$, note that $a^*_{\mathrm{I},\lambda_1} \ge b^*_{\sup}$ and $a^*_{\mathrm{I},\lambda_2} \ge b^*_{\sup}$. It is easy to see that:

$$\forall a \ge b^*_{\sup},\quad \frac{\mathrm{d}\Phi_\mathrm{I}(a)}{\mathrm{d}a} \ge 0. \tag{A7}$$

Contrary to the conclusion, we suppose that $a^*_{\mathrm{I},\lambda_1} < a^*_{\mathrm{I},\lambda_2}$. Then $\Phi_\mathrm{I}(a^*_{\mathrm{I},\lambda_1}) < \Phi_\mathrm{I}(a^*_{\mathrm{I},\lambda_2})$. Further, from:

$$\left(\Phi_\mathrm{I}\left(a^*_{\mathrm{I},\lambda_2}\right)\right)^2 + \lambda_2 V_\mathrm{I}\left(a^*_{\mathrm{I},\lambda_2}\right) \le \left(\Phi_\mathrm{I}\left(a^*_{\mathrm{I},\lambda_1}\right)\right)^2 + \lambda_2 V_\mathrm{I}\left(a^*_{\mathrm{I},\lambda_1}\right),$$

we have $V_\mathrm{I}(a^*_{\mathrm{I},\lambda_2}) < V_\mathrm{I}(a^*_{\mathrm{I},\lambda_1})$. Thus:

$$\begin{aligned}0 &\le \left(\Phi_\mathrm{I}\left(a^*_{\mathrm{I},\lambda_1}\right)\right)^2 + \lambda_2 V_\mathrm{I}\left(a^*_{\mathrm{I},\lambda_1}\right) - \left(\left(\Phi_\mathrm{I}\left(a^*_{\mathrm{I},\lambda_2}\right)\right)^2 + \lambda_2 V_\mathrm{I}\left(a^*_{\mathrm{I},\lambda_2}\right)\right)\\ &= \left(\Phi_\mathrm{I}\left(a^*_{\mathrm{I},\lambda_1}\right)\right)^2 - \left(\Phi_\mathrm{I}\left(a^*_{\mathrm{I},\lambda_2}\right)\right)^2 + \lambda_2\left(V_\mathrm{I}\left(a^*_{\mathrm{I},\lambda_1}\right) - V_\mathrm{I}\left(a^*_{\mathrm{I},\lambda_2}\right)\right)\\ &< \left(\Phi_\mathrm{I}\left(a^*_{\mathrm{I},\lambda_1}\right)\right)^2 - \left(\Phi_\mathrm{I}\left(a^*_{\mathrm{I},\lambda_2}\right)\right)^2 + \lambda_1\left(V_\mathrm{I}\left(a^*_{\mathrm{I},\lambda_1}\right) - V_\mathrm{I}\left(a^*_{\mathrm{I},\lambda_2}\right)\right).\end{aligned}$$

That is, $(\Phi_\mathrm{I}(a^*_{\mathrm{I},\lambda_1}))^2 + \lambda_1 V_\mathrm{I}(a^*_{\mathrm{I},\lambda_1}) > (\Phi_\mathrm{I}(a^*_{\mathrm{I},\lambda_2}))^2 + \lambda_1 V_\mathrm{I}(a^*_{\mathrm{I},\lambda_2})$, which contradicts with the fact that $(\Phi_\mathrm{I}(a))^2 + \lambda_1 V_\mathrm{I}(a)$ achieves a minimum at $a^*_{\mathrm{I},\lambda_1}$. Therefore, it can be concluded that $a^*_{\mathrm{I},\lambda_1} \ge a^*_{\mathrm{I},\lambda_2}$. ■

Appendix 3: Proof of result 6

From assumption 10:

$$\begin{aligned}\frac{\sum_{k=1}^{N-1}\eta_k + \gamma}{\sum_{k=1}^{N}\mu_k + \sum_{k=1}^{N-1}\eta_k + \gamma} \le 50\% \Rightarrow \Phi_\mathrm{II}(N)\\ \le \frac{\max(c_\mathrm{m}, c_\mathrm{r})}{2},\ \forall N.\end{aligned} \tag{A8}$$

When $c_\mathrm{m} > c_\mathrm{r}$:

$$\begin{aligned}\Psi_\mathrm{II}(N+1) - \Psi_\mathrm{II}(N) &= \frac{\left\{c_\mathrm{m}^2 f_N - c_\mathrm{r}^2\gamma(\mu_{N+1} + \eta_N)\right\}}{\Delta_N}\\ &> \frac{c_\mathrm{m}(c_\mathrm{m} f_N - c_\mathrm{r}\gamma(\mu_{N+1} + \eta_N))}{\Delta_N}\\ &= c_\mathrm{m}(\Phi_\mathrm{II}(N+1) - \Phi_\mathrm{II}(N)).\end{aligned}$$

Based on result 5:

$$\begin{aligned}&(\Phi_\mathrm{II}(N+1))^2 + \lambda V_\mathrm{II}(N+1) - [(\Phi_\mathrm{II}(N))^2 + \lambda V_\mathrm{II}(N)]\\ &\quad = ((1-\lambda)(\Phi_\mathrm{II}(N) + \Phi_\mathrm{II}(N+1)))\\ &\qquad \times (\Phi_\mathrm{II}(N+1) - \Phi_\mathrm{II}(N)) + \lambda(\Psi_\mathrm{II}(N+1) - \Psi_\mathrm{II}(N))\\ &\quad > (\Phi_\mathrm{II}(N) + \Phi_\mathrm{II}(N+1) - \lambda(\Phi_\mathrm{II}(N)\\ &\qquad + \Phi_\mathrm{II}(N+1) - c_\mathrm{m}))(\Phi_\mathrm{II}(N+1) - \Phi_\mathrm{II}(N)).\end{aligned}$$

From Equation (A8), $\Phi_\mathrm{II}(N) + \Phi_\mathrm{II}(N+1) - \lambda(\Phi_\mathrm{II}(N) + \Phi_\mathrm{II}(N+1) - c_\mathrm{m}) > 0$. Also, it is easy to see that if $c^*_{\inf} < \infty$, $\forall N \ge c^*_{\inf}$, then $\Phi_\mathrm{II}(N+1) - \Phi_\mathrm{II}(N) \ge 0$. Hence, $\forall N \ge c^*_{\inf}$, we have:

$$(\Phi_\mathrm{II}(N+1))^2 + \lambda V_\mathrm{II}(N+1) - [(\Phi_\mathrm{II}(N))^2 + \lambda V_\mathrm{II}(N)] > 0,$$

from which it follows that $N^* \le c^*_{\inf}$.

To prove that $N^*_{\lambda_1} \leq N^*_{\lambda_2}$ with $\lambda_1 > \lambda_2 \geq 0$, note that $N^*_{\lambda_1} \leq c^*_{\inf}$ and $N^*_{\lambda_2} \leq c^*_{\inf}$. It is easy to see that:

$$\forall N \leq c^*_{\inf},\ \Phi_{\rm II}(N) - \Phi_{\rm II}(N-1) \leq 0. \tag{A9}$$

Contrary to the conclusion, we suppose that $N^*_{\lambda_1} > N^*_{\lambda_2}$. Then, $\Phi(N^*_{\lambda_1}) < \Phi(N^*_{\lambda_2})$. Further, from:

$$\left(\Phi_{\rm II}\left(N^*_{\lambda_2}\right)\right)^2 + \lambda_2 V_{\rm II}\left(N^*_{\lambda_2}\right) \leq \left(\Phi_{\rm II}\left(N^*_{\lambda_1}\right)\right)^2 + \lambda_2 V_{\rm II}\left(N^*_{\lambda_1}\right),$$

we have $V_{\rm II}(N^*_{\lambda_2}) < V_{\rm II}(N^*_{\lambda_1})$. Thus:

$$\begin{aligned}
0 &\leq \left(\Phi_{\rm II}\left(a^*_{1,\lambda_1}\right)\right)^2 + \lambda_2 V_{\rm II}\left(a^*_{1,\lambda_1}\right) - \left(\left(\Phi_{\rm II}\left(N^*_{\lambda_2}\right)\right)^2 \right.\\
&\quad \left. + \lambda_2 V_{\rm II}\left(N^*_{\lambda_2}\right)\right)\\
&= \left(\Phi_{\rm II}\left(N^*_{\lambda_1}\right)\right)^2 - \left(\Phi_{\rm II}\left(N^*_{\lambda_2}\right)\right)^2 + \lambda_2\left(V_{\rm II}\left(N^*_{\lambda_1}\right) - V_{\rm II}\left(N^*_{\lambda_2}\right)\right)\\
&< \left(\Phi_{\rm II}\left(N^*_{\lambda_1}\right)\right)^2 - \left(\Phi_{\rm II}\left(N^*_{\lambda_2}\right)\right)^2 + \lambda_1\left(V_{\rm II}\left(N^*_{\lambda_1}\right) - V_{\rm II}\left(N^*_{\lambda_2}\right)\right).
\end{aligned}$$

That is, $(\Phi_{\rm II}(N^*_{\lambda_1}))^2 + \lambda_1 V_{\rm II}(N^*_{\lambda_1}) > (\Phi_{\rm II}(N^*_{\lambda_2}))^2 + \lambda_1 V_{\rm II}(N^*_{\lambda_2})$, which contradicts with the fact that $(\Phi_{\rm II}(N))^2 + \lambda_1 V_{\rm II}(N)$ achieves a minimum at $N^*_{\lambda_1}$. Therefore, it can be concluded that $N^*_{\lambda_1} \leq N^*_{\lambda_2}$.

When $c_{\rm m} < c_{\rm r}$:

$$\begin{aligned}
\Psi_{\rm II}(N) - \Psi_{\rm II}(N-1) &= \frac{\left\{c^2_{\rm m} f_{N-1} - c^2_{\rm r}\gamma(\mu_N + \eta_{N-1})\right\}}{\Delta_{N-1}}\\
&< \frac{c_{\rm r}(c_{\rm m} f_{N-1} - c_{\rm r}\gamma(\mu_N + \eta_{N-1}))}{\Delta_{N-1}}\\
&= c_{\rm r}(\Phi_{\rm II}(N) - \Phi_{\rm II}(N-1)).
\end{aligned}$$

Based on result 5:

$$\begin{aligned}
&(\Phi_{\rm II}(N))^2 + \lambda V_{\rm II}(N) - [(\Phi_{\rm II}(N-1))^2 + \lambda V_{\rm II}(N-1)]\\
&\quad = ((1-\lambda)(\Phi_{\rm II}(N-1) + \Phi_{\rm II}(N)))(\Phi_{\rm II}(N) - \Phi_{\rm II}(N-1))\\
&\qquad + \lambda(\Psi_{\rm II}(N) - \Psi_{\rm II}(N-1))\\
&\quad < (\Phi_{\rm II}(N-1) + \Phi_{\rm II}(N) - \lambda(\Phi_{\rm II}(N-1)\\
&\qquad + \Phi_{\rm II}(N) - c_{\rm m}))(\Phi_{\rm II}(N) - \Phi_{\rm II}(N-1)).
\end{aligned}$$

From Equation (A8), $\Phi_{\rm II}(N-1) + \Phi_{\rm II}(N) - \lambda(\Phi_{\rm II}(N-1) + \Phi_{\rm II}(N) - c_{\rm r}) > 0$. Also, it is easy to see that $\forall N \leq c^*_{\sup}$ we have $\Phi_{\rm II}(N) - \Phi_{\rm II}(N-1) \leq 0$. Hence, $\forall N \leq c^*_{\sup}$:

$$(\Phi_{\rm II}(N))^2 + \lambda V_{\rm II}(N) - [(\Phi_{\rm II}(N-1))^2 + \lambda V_{\rm II}(N-1)] < 0,$$

from which it follows that $N^* \geq c^*_{\sup}$.

To prove $N^*_{\lambda_1} \geq N^*_{\lambda_2}$ with $\lambda_1 > \lambda_2 \geq 0$, note that $N^*_{\lambda_1} \geq c^*_{\sup}$ and $N^*_{\lambda_2} \geq c^*_{\sup}$. It is easy to see that:

$$\forall N \geq c^*_{\sup},\quad \Phi_{\rm II}(N+1) - \Phi_{\rm II}(N) \geq 0. \tag{A10}$$

Contrary to the conclusion, we suppose that $N^*_{\lambda_1} < N^*_{\lambda_2}$. Then, $\Phi_{\rm II}(N^*_{\lambda_1}) < \Phi_{\rm II}(N^*_{\lambda_2})$. Further, from:

$$\left(\Phi_{\rm II}\left(N^*_{\lambda_2}\right)\right)^2 + \lambda_2 V_{\rm II}\left(N^*_{\lambda_2}\right) \leq \left(\Phi_{\rm II}\left(N^*_{\lambda_1}\right)\right)^2 + \lambda_2 V_{\rm II}\left(N^*_{\lambda_1}\right),$$

we have $V_{\rm II}(N^*_{\lambda_2}) < V_{\rm II}(N^*_{\lambda_1})$. Thus:

$$\begin{aligned}
0 &\leq \left(\Phi_{\rm II}\left(N^*_{\lambda_1}\right)\right)^2 + \lambda_2 V_{\rm II}\left(N^*_{\lambda_1}\right) - \left(\left(\Phi_{\rm II}\left(N^*_{\lambda_2}\right)\right)^2 + \lambda_2 V_{\rm II}\left(N^*_{\lambda_2}\right)\right)\\
&= \left(\Phi_{\rm II}\left(N^*_{\lambda_1}\right)\right)^2 - \left(\Phi_{\rm II}\left(N^*_{\lambda_2}\right)\right)^2 + \lambda_2\left(V_{\rm II}\left(N^*_{\lambda_1}\right) - V_{\rm II}\left(N^*_{\lambda_2}\right)\right)\\
&< \left(\Phi_{\rm II}\left(N^*_{\lambda_1}\right)\right)^2 - \left(\Phi_{\rm II}\left(N^*_{\lambda_2}\right)\right)^2 + \lambda_1\left(V_{\rm II}\left(N^*_{\lambda_1}\right) - V_{\rm II}\left(N^*_{\lambda_2}\right)\right).
\end{aligned}$$

That is, $(\Phi_{\rm II}(N^*_{\lambda_1}))^2 + \lambda_1 V_{\rm II}(N^*_{\lambda_1}) > (\Phi_{\rm II}(N^*_{\lambda_2}))^2 + \lambda_1 V_{\rm II}(N^*_{\lambda_2})$, which contradicts with the fact that $(\Phi_{\rm II}(N))^2 + \lambda_1 V_{\rm II}(N)$ achieves a minimum at $N^*_{\lambda_1}$. Therefore, it can be concluded that $N^*_{\lambda_1} \geq N^*_{\lambda_2}$. ■

Biographies

Yong Chen is currently an assistant professor in the Department of Mechanical and Industrial Engineering at the University of Iowa. He received the B.E. degree in computer science from Tsinghua University, China in 1998, the Master degree in Statistics and Ph.D. degree in Industrial & Operations Engineering from the University of Michigan in 2003. His research interests include the fusion of advanced statistics, applied operations research, and engineering knowledge to develop in-process quality and productivity improvement methodologies achieving automatic fault diagnosis, proactive maintenance, and integration of quality and reliability in complex manufacturing systems. He is a member of ASQ, INFORMS, and IIE.

Jionghua (Judy) Jin received her B.S. and M.S. degrees in Mechanical Engineering, both from Southeast University in 1984 and 1987 respectively, and her Ph.D. degree in Industrial and Operations Engineering at the University of Michigan in 1999. She is currently an Assistant Professor in the Department of Systems and Industrial Engineering at the University of Arizona. Her research focuses on developing a unified methodology for quality and reliability improvement through the fusion of statistics methods with engineering models. Her research expertise is in the areas of systematic process modeling for variation analysis, automatic feature extraction for monitoring and fault diagnosis, and optimal maintenance decision with the integration of the quality and reliability interaction. Her research has been applied to various complex systems such as multistage manufacturing processes including assembly, stamping, semiconductor manufacturing, and the transportation service industry. Her current research is being sponsored by the National Science Foundation, SME Education Foundation, the US Department of Transportation's Bureau of Transportation Statistics, Arizona State Foundation, and Global Solar Energy Inc. She currently serves on the Editorial Board of *IIE Transactions on Quality and Reliability*. She is a member of INFORMS, IIE, ASQC, ASME, and SME. She was the recipient of a CAREER Award from the National Science Foundation in 2002, and the Best Paper Award from ASME, Manufacturing Engineering Division in 2000.

Contributed by the Reliability Engineering Department