Jointly optimizing lot sizing and maintenance policy for a production system with two failure modes[1]

Kaiye Gao[a,b,*], Rui Peng[c], Li Qu[a], Shaomin Wu[d]

[a]School of Economics and Management, Beijing Information Science and Technology University, Beijing 100192 China

[b]Academy of Mathematics and System Science, Chinese Academy of Sciences, Beijing 100190, China

[c]School of Economics and Management, Beijing University of Technology, Beijing 100124 China

[d]Kent Business School, University of Kent, Canterbury CT2 7FS England

Email and address

Kaiye Gao (*Corresponding author):

kygao@foxmail.com, Anxiangnanli, Chaoyang District, Beijing, 100101, China

Rui Peng:

pengruisubmit@163.com, 100 Pingleyuan, Chaoyang District, Beijing, 100124, China

Li Qu:

quli@bistu.edu.cn, 12 Xiaoyingdong Road, Haidian District, Beijing, 100192, China

Shaomin Wu:

s.m.wu@kent.ac.uk, Kent Business School, University of Kent, Canterbury, CT2 7FS, UK

[1] Suggested citation: Kaiye Gao, Rui Peng, Li Qu, Shaomin Wu, Jointly optimizing lot sizing and maintenance policy for a production system with two failure modes, *Reliability Engineering & System Safety*, DOI: 10.1016/j.ress.2020.106996

Abstract: In the reliability literature, there are studies that jointly study maintenance and production and that is typically restricted to one failure mode, and fail to address the case where multiple failure modes exist. This study investigates the problem of joint optimization of lot sizing and maintenance policy for a multi-product production system subject to two failure modes. The failure of the first mode refers to the soft failure that occurs after defects arrive. The failure of the second mode is a hard failure that occurs without any early warning signals. Products are sequentially produced by the system and a complete run of all products forms a production cycle. The system needs to be re-set up before producing a different product. Both the production cycle and the set-up point depend on the lot sizes of products. Models are proposed for two maintenance policies: 1) arranging the maintenance to be at the end of each production cycle; 2) arranging the maintenance to be at set-up points. The expected profit per unit time is formulated to obtain the optimal lot sizing and maintenance policy. Some properties of proposed models are proved, which show that the optimal lot sizing and maintenance policy can be obtained under certain conditions. Case studies and sensitivity analyses are presented to illustrate the proposed models of two maintenance policies. Basically, the results show that the producer will gain the most profit if the optimal lot sizing and maintenance policy are adopted. The results of comparing both maintenance policies reveal that the excessive maintenance is not economic. The sensitivity analyses illustrate that reducing the cost caused by failures and improving system reliability are effective ways to increase the expected profit per unit time.

Keywords: Maintenance; Production; Failure modes; Lot sizing; Reliability

1. Introduction

In the manufacturing industry, small lots of items may be produced so that inventory cost can be reduced, which, however, may incidentally increase set-up cost. This raises a challenge on how the lot size can be optimised to improve the economic benefit of the manufacturer, which has been widely studied for a long time (Ben-Daya et al., 2008; Lamas & Chevalier, 2018; Kilic & Tunc, 2019; Ou & Feng, 2019; Taş et al., 2019). In most cases, the optimal lot sizing is obtained by minimising the sum of inventory holding and set-up costs (Liu et al., 2019). In practice, a production system may fail and maintenance may therefore be carried out (Barata et al., 2002; Garbatov & Soares, 2001;Xiao & Peng, 2014; Levitin & Lisnianski, 2000; Cha et al., 2017; Cha et al., 2018; Yang, et al., 2018; Yang et al., 2019; Zhang et al., 2014; Li et al., 2016). This raises another challenge on how the maintenance can be optimally scheduled to reduce the possibility of system's future failures.

Rising production costs and efficiency requirements are challenges for manufacturers (Schreiber et al., 2019). One important way to cope with these challenges is to improve maintenance effectiveness, on which there are many publications (Rivera-Gómez et al., 2020; Zhou & Yu 2020; Nguyen, 2019; Zhou et al., 2018; Cheng et al., 2018). For a multi-product system, it may need to be re-set up before producing a new lot of product items. In this case, the set-up epoch can be utilized as a maintenance window to reduce the interruption caused by the maintenance actions.

Herein, some studies optimized maintenance policies in conjunction with lot-size determination (Liu et al., 2015; Lu et al., 2013; Ben-Daya & Noman, 2006). Particularly, Ben-Daya & Noman (2006) developed an integrated model that considers simultaneously inventory production decisions, preventive maintenance (PM) schedule, and warranty policy. With this integrated model, it is illustrated through numerical examples that investment in PM can lead to savings in warranty claims for repairable products. As a result, the overall profit per unit, in certain cases, may be higher with PM than without PM. Lu et al. (2013) proposed a joint model for integrating run-based preventive maintenance (PM) into the capacitated lot sizing problem. They assumed that both production and PM operations are restricted by the system's maximum capacity and that the system reliability has to be maintained above a threshold value throughout the planning horizon. Liu et al. (2015) constructed an integrated production, inventory and preventive maintenance model. They concluded that the product lot sizes and the PM policy should be jointly optimised since they influence each other in terms of cost and profit.

All the above-mentioned studies are restricted to a limitation that there is only one failure mode in such systems, which is normally not the case in practice, especially given that fact that a modern production system is composed of many components. For example, Ben-Daya & Noman (2006) presented a deteriorating system that experiences shifts to an out of control state. Lu et al. (2013) assumed that the production system is subject to deterioration with usage. Liu et al. (2015) just considered the case where the failure of components is a two-stage process. However, many systems may fail due to both hard failures and soft failures in practice (Peng et al., 2019; Peng et al., 2010; Wang & Wu, 2014). Hard failures are self-announcing failures with instantaneous occurrence (Ye et al., 2014), while soft failures are failures with early warning signals, which can be detected by inspection or monitoring (Zhao et al., 2015). Modern production systems usually consist of both mechanical and electrical components. Mechanical components usually suffer soft failures such as wears whose failure may only be exposed by inspection, whereas electronic components usually suffer hard failure and may fail without any early warning. Motivated by this fact, this paper jointly determine the optimal lot sizing and maintenance policy for a production system that produces multiple products considering two different failure modes: failure modes I and II. Failure mode I represents the soft failure whose failure process is characterized with the delay-time concept, i.e., the failure process consists of two stages, the normal stage from normal to defective and the delay time stage from defective to failure, as shown in **Fig. 1** (Mahmodi et al., 2017; Christer & Wang, 1995). Failure mode II is the hard failure whose failure happens without early symptoms and the failure rate increases with operating time. PM are conducted to detect and repair the possible defects of failure mode I in the system (Wu & Zuo, 2010; Wu & Clements-Croome, 2005; Cheng et al., 2018). In order to prevent the system from unexpected interruptions due to failure mode II, overhaul (OH) is implemented to restore the system to be the good as new status. In order to reduce the interruption to the production process, it is assumed that PM and OH are scheduled to be at the set-up epoch for different products during the production. Herein, the maintenance policy depends on the lot sizing and thus, the objective of this paper is to maximise the expected profit while jointly optimizing the lot sizing and the maintenance policy.

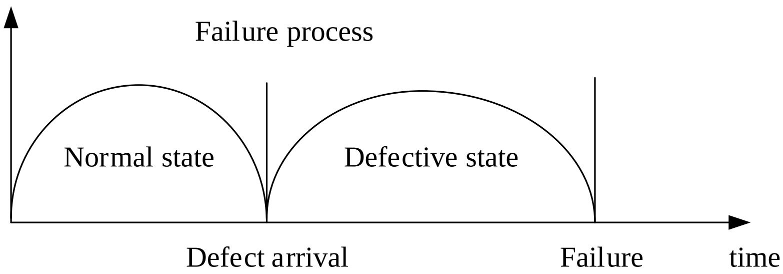


Fig. 1. Delay time concept.

This study extends two prior papers: Liu et al. (2015) and Peng et al. (2019). The former considers joint optimisation of maintenance policy and production planning for the production system having only soft failure, whereas the latter considers merely the maintenance policy and does not consider production planning. Jointly considering the two types of failure in optimisation of maintenance policy and production strategy is a challenging task for three reasons. First, the two failure modes are different in terms of what/which increases the complexity in analyzing the failure process of the system. Second, multiple maintenance actions for different failure modes are used and therefore their combinations. Last but not least, because there are multiple decision variables regarding lot sizing and maintenance policy for this optimization problem, the analytic solution is difficult to be obtained. These challenges require us to model more complex situations and to propose more complex mathematical propositions.

The remainder of this paper is arranged as follows. Section 2 gives system description and assumptions. Section 3 introduces a model for the case where the system is maintained at the end of the production cycle. Section 4 extends the model to the case where the system is maintained at the set-up point. Section 5 studies the properties of optimal policy and an approach to obtaining it. Section 6 presents case studies and sensitivity analyses to illustrate the applications. Section 7 concludes this study and suggests future research.

Nomenclature

Abbreviations and Acronyms

PM	Preventive maintenance
OH	Overhaul
MR	Minimal repair
HPP	Homogeneous Poisson process
pdf	Probability density function
cdf	Cumulative density function

Notations

k	Total different number of product types
i	Index of the type of product

D_i	Total demand of the ith product during the period of concern
d_i	Total consumption rate of the ith product
p_i	Total production rate of the ith product
n	Total number of the production cycles in the study period
T_i	Nominal production time of the ith product in one production cycle
S	Number of PMs until a following OH is carried out
u	Time point of a defect arrival
δ	Rate of the defect occurrence
t_1	Delay-time of defect in failure mode I
$f_1(t_1)$	pdf of t_1
$F_1(t_1)$	cdf of t_1
t_2	Time of mode II failure
$f_2(t_2)$	pdf of t_2
$F_2(t_2)$	cdf of t_2
$\lambda(t_2)$	Hazard rate of t_2
C_h^i	Average inventory holding cost per unit product per unit time for the ith product
C_s^i	Average cost per set-up for the ith product
C_d	Average cost of repairing a defective component including the additional cost due to unavailability
C_p	Average cost of an inspection at PM
C_o	Average cost of an OH
C_f^*	Average cost of repairing a failure, which contains the cost for repairing or replacing the failed component and the additional cost of unavailability, where $*=1$ and 2, indicating failure modes I and II respectively
C_F^*	Total cost of repairing failures within the period of concern, where $*=1$ and 2 indicating failure modes I and II respectively
R_i	The gross profit per unit product i, which is equal to the unit sale price minus the unit production costs excluding the maintenance, set-up and inventory costs
$E(C;*)$	Expected cost of set-up (s), holding (h), and maintenance (m) within the period of concern

	where $* = s, h$ and m, respectively
$E(C;m;*)$	Expected cost of OH (*o*), PM (*p*), and failure (*f*) which constitute the expected maintenance cost within the period of concern where $* = o, p$ and f, respectively
$E(C;n,S)$	Expected total cost of the period of concern with n and S as the decision variables
$E(P;n,S)$	Expected profit per unit time with n and S as the decision variables

2. System Description and Assumptions

2.1. System description

In this study, we consider a system that produces multiple products. The fixed demands of different products within a certain period are divided into small lot sizes and the products are produced in sequence. Here a production cycle is a complete run of all products during the production process, as shown in **Fig.2**. Each product is produced exactly once within a production cycle and the cycle is repeated over time. Each time when the system turns to produce a different product, the system needs to be re-set up. These set-up time epochs can be utilized for maintenance to avoid interrupting the production process.

The production system under consideration is subjected to two different and statistically independent failure modes. Failure mode I is modeled based on the delay-time concept and its failure process includes two stages: a normal stage and a defective stage. The defective stage can be detected by inspection. For a multi-component system, the arrival of defects is assumed to follow a homogeneous Poisson process (HPP) (Christer & Wang, 1995). In contrast, failure mode II corresponds to the failure occurrence without any early warning.

Three types of maintenance, PM, OH and minimal repair (MR), are applied. PM is used to inspect possible defects and then fix them. PM is assumed to be able to fix all the defects due to failure mode I but it does not affect the failure rate of mode II. OH, however, can fix defects and further restore the system to be good as new. In particular, OH can not only fix defects due to mode I but also reduce the failure rate of mode II to the level as it was at the beginning of the production. In case where the system fails between two successive preventive actions (PM or OH), MR is used to restore the system to work without changing the occurrence process of either failure mode.

For the sake of practical implementation convenience, we assume that an OH is carried out after every S consecutive PMs. Hence, we introduce the concept of maintenance cycle and define it as a complete run of S PMs and one OH. Two different maintenance policies are studied, where the first policy arranges PMs and OHs to be at the end of the production cycles and the second policy arranges PMs and OHs to be at set-up periods. The production process under consideration are under infinite time horizon, and the excepted total profit per unit time is used as the objective to be maximized, where n (the total number of the production cycles) and S are two decision variables. The models under the two policies are constructed in Section 3 and Section 4, which are referred to as "Model of Case 1" and "Model of Case 2", respectively.

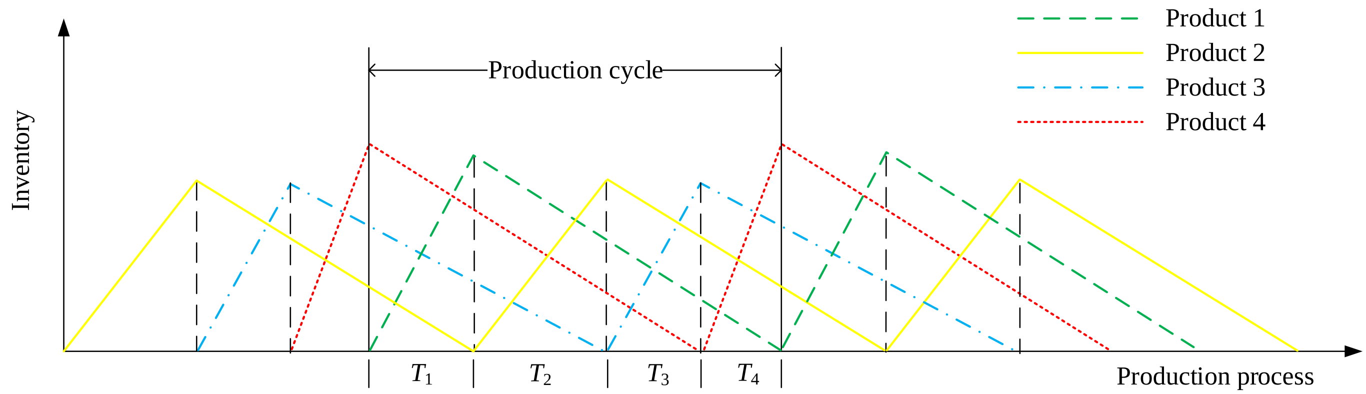


Fig. 2. A typical example of a production cycle with four products.

2.2. Assumptions

1) The demands of all products are fixed. For each type of product, its demand can be divided into n small lot sizes that are produced according to a preset and fixed sequence. The production process is within an infinite time horizon.

2) Each product is produced once in a production cycle, and the production cycle is a complete run of all products according to their lot sizes.

3) The production system is subject to two different, independent failure modes.

4) The failure process of mode I is divided into two stages, normal and delay time stages, where the arrival of defects follows an HPP with occurrence rate δ. The delay time of all defects is independently and identically distributed.

5) The failure rate of failure mode II increases with the operating time and the failure of this mode occurs without early warning symptoms.

6) PMs and OHs are carried out at set-up points to reduce interruption to the production. A PM only detects and fixes all defects with respect to failure mode I, whereas OH renews the whole system to prevent both two failure modes.

7) An OH is carried out after every S consecutive PMs. S consecutive PMs and a following OH constitutes a maintenance cycle.

8) When a failure occurs, a minimal repair is always performed. The minimal repair resumes the system from the failure without changing the failure process.

9) Maintenance time on PM, OH, set-up and failures are negligible.

Assumption (1) is directly abstracted from the lot production when its demand is fixed. Assumptions (2) and (3) have already been explained in Section 2.1. Assumption (4) was used in previous studies based on delay-time models (Liu et al., 2013). However, it should be noted that even though the defect arrival follows an HPP, the failure process of the defect is a Non-Homogeneous Poisson Process (NHPP). Assumption (5) is an approximation from the industry practice, because failure mode II is an outcome of many unobserved and unpredictable factors so that the failure rate and the failure process due to this mode can only be described with operating time. The failure rate is assumed to increase as time, as most systems are degrading with time. For example, the failure time is usually described by the Weibull distribution with shape parameter bigger than 1. Assumption (6) is a fact observed in the industry of typical lot production: the set-up epoch is usually used for PM and OH in case of system failure. This is particularly true for continuous lot production, such as steel production, where the interruption cost of production caused by maintenance or failure is high. Assumption (7) is mathematically an expression for the maintenance cycle according to assumption (6) and system description. Compared with PM, OH is a thorough repair, so it is usually more costly. Herein, OH should be conducted less frequently than PM. Assumption (8) is widely used in maintenance modeling (Liao, 2012). In the industry of typical continuous lot production, resuming the production from its failure is usually urgent so that the repair time is limited. In this case, the most cost-effective way is to detect and further repair or replace only the failed component, which is called minimal repair. As a result, this repair mode basically cannot influence the overall system defect rate and failure intensities. Assumption (9) is an approximation which is used to simplify the modeling process. In fact, compared with the production time (usually calculated by month or year), the downtimes caused by PM, OH, set-up and system failure(usually calculated by hour) are much shorter.

3. Model of Case 1: maintaining the system at the end of each production cycle

In this case, a maintenance cycle contains $S+1$ consecutive production cycles. For a maintenance cycle, PM is carried out at the end of each of the first S production cycles, whereas OH is carried out at the end of the $(S+1)$ th production cycle. Since the maintenance cycle repeats during production in terms of both lot sizing and maintenance policy, we only need to consider one maintenance cycle as the study period. The duration of a production cycle is $\sum_{i=1}^{k} T_i$, where T_i is the nominal production time of the *i*th product in one production cycle. Accordingly, the duration of a maintenance cycle is $(S+1)\sum_{i=1}^{k} T_i$.

To calculate the gross profit, we first estimate the total cost of a maintenance cycle. This total consists of maintenance costs, inventory holding cost and set-up cost. Then the expected total profit of a maintenance cycle can be obtained by subtracting the total cost from the total gross profit. Finally, the expected profit per unit time is calculated by dividing the total profit with the duration of a maintenance cycle.

3.1. Maintenance costs

The maintenance costs incur due to OH, PM and failure repair. In a maintenance cycle, OH is only carried out once, so the OH cost is $E(C;m;o)=C_o$. Now we model the PM cost and failure repair cost in a maintenance cycle, which are shown as follows.

3.1.1. PM cost

The cost of PM has two parts: inspection and defect repair. Assume that the average cost of a PM is C_p, then the cost of inspection for all PMs is $C_p S$.

To calculate the repair cost of defects, we need to determine the expected number of defects presented at PM. Consider the event that a defect arising in $(u, u+du)$ still exists at time t. As the arrival of defects follows an HPP with occurrence rate δ, the probability of this event is

$$\delta du P\left(h > t-u\right) = \delta\left[1-\int_0^{t-u} f_1\left(h\right)dh\right]du. \quad (1)$$

Thus, the expected number of defects identified by the inspection at a PM is the number of defects that arise in a production cycle and still present at the end of this cycle, that is $\int_0^{\sum_{i=1}^k T_i} \delta\left[1-\int_0^{\sum_{i=1}^k T_i - t_1} f_1\left(h\right)dh\right]dt_1 = \int_0^{\sum_{i=1}^k T_i} \delta\left[1-F_1\left(t_1\right)\right]dt_1$.

Then, the cost of fixing defects during a maintenance cycle is $C_d S\int_0^{\sum_{i=1}^k T_i} \delta\left[1-F_1\left(t_1\right)\right]dt_1$. It should be noted that the defects arising in the last production cycle in a maintenance cycle are directly removed by the OH at the end of the maintenance cycle. By adding up the two parts together, we have the expected PM cost as

$$E(C;m;p) = C_d S\int_0^{\sum_{i=1}^k T_i} \delta\left[1-F_1\left(t_1\right)\right]dt_1 + C_p S. \quad (2)$$

3.1.2. Repair cost

Repair cost incurs due to two failure modes. For mode I, the probability that a defect arriving within $(u, u+du)$ leads to a failure before time t is

$$\delta du P\left(h = t-u\right) = \delta du\int_0^{t-u} f_1\left(h\right)dh. \quad (3)$$

Then, the expected number of failures caused by defects in a production cycle is $\int_0^{\sum_{i=1}^k T_i} \delta\int_0^{\sum_{i=1}^k T_i - t_1} f_1\left(h\right)dh dt_1 = \int_0^{\sum_{i=1}^k T_i} \delta F_1\left(t_1\right)dt_1$. If the cost of repairing a failure of mode I is C_f^1, we obtain the expected cost due to repairing failures of mode I

$$C_F^1 = C_f^1\left(S+1\right)\int_0^{\sum_{i=1}^k T_i} \delta F_1\left(t_1\right)dt_1. \quad (4)$$

For failure mode II, the pdf ofthe failure occurrence at time t for a renewed system is $f_2\left(t\right)$. Since the minimal repair does not change the failure process of failure mode II, the expected number of failures during $(u, u+du)$ is $\lambda_2\left(u\right)du$, where $\lambda_2\left(u\right) = f_2\left(u\right)/\left(1-F_2\left(u\right)\right)$ is the hazard rate. Thus, the expected number of failures in a maintenance cycle is

$\int_0^{(S+1)\sum_{i=1}^k T_i} \lambda_2(t_2)dt_2$. Then if the cost of repairing a failure due to mode II is C_f^2 , we obtain the expected cost due to repairing failures of mode II

$$C_F^2 = C_f^2 \int_0^{(S+1)\sum_{i=1}^k T_i} \lambda_2(t_2)dt_2. \tag{5}$$

Finally, by adding up the repair costs from the two failure modes, the expected total cost of repairing failure in a maintenance cycle can be obtained, as

$$E(C;m;f) = C_F^1 + C_F^2 = C_f^1 (S+1) \int_0^{\sum_{i=1}^k T_i} \delta F_1(t_1)dt_1 + C_f^2 \int_0^{(S+1)\sum_{i=1}^k T_i} \lambda_2(t_2)dt_2. \tag{6}$$

3.1.3. Total maintenance costs

Now we have the expected, total maintenance costs in a maintenance cycle which is the sum of OH cost, PM cost and failure repair cost, as

$$\begin{aligned} E(C;m) &= E(C;m;o) + E(C;m;p) + E(C;m;f) \\ &= C_o + C_d S \int_0^{\sum_{i=1}^k T_i} \delta[1 - F_1(t_1)]dt_1 + C_p S + C_f^1 (S+1) \int_0^{\sum_{i=1}^k T_i} \delta F_1(t_1)dt_1 + C_f^2 \int_0^{(S+1)\sum_{i=1}^k T_i} \lambda_2(t_2)dt_2. \end{aligned} \tag{7}$$

3.2. Inventory holding cost

To estimate the inventory holding cost during a maintenance cycle, firstly we need to quantify the inventory of a production cycle, since the production, consumption and inventory are the same in each production cycle.

According to the notation and assumptions, D_i is the total demand of the *i*th product and n is the number of production cycles. Then, we know that the lot size of product i for each production cycle is D_i / n and the consumption rate is $d_i = D_i / \left(n \sum_{i=1}^k T_i\right)$.

Now the maximum inventory quantity of product i can be obtained by multiplying the difference between the production rate and the consumption rate with the production time, which is $(p_i - d_i)T_i$. If both the production rate and the consumption rate are fixed, the total inventory holding of product i in a production cycle is $(1/2)\left(\sum_{i=1}^k T_i\right)(p_i - d_i)T_i$. Then, we have the expected inventory holding cost in a maintenance cycle, which is

$$E(C;h) = (S+1)\left(\sum_{i=1}^k T_i\right)\left[\sum_{j=1}^k (p_j - d_j) T_j C_h^j\right] \Big/ 2 = (S+1)\left(\sum_{i=1}^k T_i\right) \times \left[\sum_{j=1}^k \left(p_j - D_j \Big/ \left(n \sum_{i=1}^k T_i\right)\right) T_j C_h^j\right] \Big/ 2. \tag{8}$$

3.3. Set-up cost

It is obvious that the set-up cost of a production cycle is $\sum_{i=1}^k C_s^i$. Then, we have the set-up cost of a maintenance cycle

$$E\left(C;s\right)=\left(S+1\right)\sum_{i=1}^{k}C_s^i. \tag{9}$$

3.4. Expected profit per unit time

Now the expected total cost during a maintenance cycle can be calculated, which is the sum of maintenance costs, inventory holding cost and set-up cost,

$$E(C;n,S)=E(C;m)+E(C;h)+E(C;s). \tag{10}$$

Assume that R_i is the gross profit per unit product i, which is equal to the unit sale price subtracting the unit production costs that exclude the maintenance, set-up and inventory costs. Then the expected total gross profit in a maintenance cycle is $\sum_{i=1}^{k}R_iD_i$. Thus, the total profit can be obtained by excluding the maintenance, set-up and inventory costs from the total gross profit, which is $\left(S+1\right)\sum_{i=1}^{k}\left(D_iR_i/n\right)-E(C;n,S)$. Finally, we have the expected profit per unit time,

$$\begin{aligned}
E(P;n,S)&=\left[\left(S+1\right)\sum_{i=1}^{k}\left(D_iR_i/n\right)-E(C;n,S)\right]\Big/\left(S+1\right)\sum_{i=1}^{k}T_i\\
&=\Bigg\{\left(S+1\right)\sum_{i=1}^{k}\left(D_iR_i/n\right)-\Bigg\{C_o+C_dS\int_0^{\sum_{i=1}^{k}T_i}\delta\left[1-F_1\left(t_1\right)\right]dt_1+C_pS+C_f^1\left(S+1\right)\int_0^{\sum_{i=1}^{k}T_i}\delta F_1\left(t_1\right)dt_1\\
&\quad+C_f^2\int_0^{(S+1)\sum_{i=1}^{k}T_i}H_2\left(t_2\right)dt_2\Bigg\}-\left\{\left(S+1\right)\left(\sum_{i=1}^{k}T_i\right)\left[\sum_{j=1}^{k}\left(p_j-D_j\Big/\left(n\sum_{i=1}^{k}T_i\right)\right)T_jC_h^j\right]\Big/2\right\}\\
&\quad-\left\{\left(S+1\right)\sum_{i=1}^{k}C_s^i\right\}\Bigg\}\Big/\left[\left(S+1\right)\sum_{i=1}^{k}T_i\right],
\end{aligned} \tag{11}$$

where $T_i=D_i/\left(np_i\right)$ is the production time of product i within a production cycle, $\sum_{i=1}^{k}T_i$ is the duration of a production cycle and $\left[\left(S+1\right)\sum_{i=1}^{k}T_i\right]$ is the duration of a maintenance cycle. One can maximize $E(P;n,S)$ to obtain the optimal n and S.

4. Model of Case 2: maintaining the system at each set-up point

In this case, a maintenance cycle contains $S+1$ consecutive set-ups, where a PM is carried out at each of the first S set-up points and an OH is carried out at $\left(S+1\right)$th set-up point. There are k set-ups in a production cycle. It should be noted that the time between two consecutive set-ups is T_i, which is the production time of product i in a production cycle.

In practice, S and k may not be an integral multiple of each other, so neither the production cycle nor the maintenance cycle forms a system renewal cycle. Thus, we need to use a combined cycle including $\left(S+1\right)k$ set-ups as the study period in this case. In this way, we can ensure that the study period is completely and repeatedly conducted during the production process. A combined cycle can be divided into $(S+1)$ production cycles or k maintenance cycles, and the duration of a combined cycle is $(S+1)\sum_{i=1}^{k} T_i$. We also start with the maintenance costs in this case.

4.1. Maintenance costs

The maintenance costs in this case are also calculated for three parts: OH, PM and failure repair. Since a combined cycle contains k maintenance cycles, the total OH cost of a combined cycle is $E(C;m;o)=C_o k$. In addition, it should be noted that OH is carried out at set-up points of fixed positions with respect to a production cycle, i.e. $\{(S+1)\%k, 2(S+1)\%k, ..., k(S+1)\%k\}$, where % is the modulo operator. Now we model the PM cost and failure repair cost in a combined cycle, which are shown as follows.

4.1.1. PM cost

The cost of PM has two parts: inspection and defect repair. The inspection cost is directly obtained according to the number of PMs in a combined cycle, which is $C_p Sk$.

According to Eq. (1), the expected total number of defects presented at set-up points during a maintenance cycle is $(S+1)\sum_{i=1}^{k}\int_0^{T_i}\delta\left[1-F_1(t_1)\right]dt_1$. Among these defects, the expected number of defects detected and fixed by OH is $\sum_{i=1}^{k}\int_0^{T_{i(S+1)\%k}}\delta\left[1-F_1(t_1)\right]dt_1$. Consequently, we obtain the expected number of defects repaired by PM. Then, we have the expected PM cost,

$$E(C;m;p)=C_p Sk + C_d\,(S+1)\sum_{i=1}^{k}\int_0^{T_i}\delta\left[1-F_1(t_1)\right]dt_1 - C_d\sum_{i=1}^{k}\int_0^{T_{i(S+1)\%k}}\delta\left[1-F_1(t_1)\right]dt_1. \quad (12)$$

4.1.2. Cost of repair upon failures

Failures are due to two failure modes. For failure mode I, according to (3), the expected number of failures occur between the ith and $(i+1)$th set-up points is $\int_0^{T_i}\delta\int_0^{t_1} f_1(t_1-u)du dt_1=\int_0^{T_i}\delta F_1(t_1)dt_1$. Thus, the repair cost for failures of mode I in a combined cycle is

$$C_F^1 = C_f^1 (S+1)\sum_{i=1}^{k}\int_0^{T_i}\delta F_1(t_1)\,dt_1. \quad (13)$$

For failure mode II, since the maintenance cycle may not synchronize with the production cycle, the time between two OHs may be different. Thus, the expected number of failures of failure mode II is changed for different maintenance cycle so that the expected number of failures in each maintenance cycle should be calculated separately. Then we have the repair cost for failures of mode II is given by

$$C_F^2 = C_f^2 \sum_{i=1}^{k} \int_0^{TF_i} \lambda_2(t_2)\,dt_2, \tag{14}$$

and,

$$\begin{aligned} TF_i = & 1\left(\lfloor i(S+1)/k \rfloor > \lfloor (i-1)(S+1)/k \rfloor\right) \\ & \times\left[\sum_{j=(i-1)(S+1)\%k+1}^{k} T_j + \sum_{j=1}^{i(S+1)\%k} T_j + \left(\lfloor i(S+1)/k \rfloor - \lfloor (i-1)(S+1)/k \rfloor - 1\right)\sum_{j=1}^{k} T_j\right] \\ & + 1\left(\lfloor i(S+1)/k \rfloor = \lfloor (i-1)(S+1)/k \rfloor\right)\sum_{j=(i-1)(S+1)\%k+1}^{i(S+1)\%k} T_j, \end{aligned} \tag{15}$$

where $1(x)=1$ if x is true; $1(x)=0$ otherwise; and $\lfloor x \rfloor$ is the integer part of x.

Finally, by adding up the repair costs for failure modes I and II, the expected total cost of repairing failures that occur in a combined cycle can obtained by

$$E(C;m;f) = C_F^1 + C_F^2. \tag{16}$$

4.1.3. Maintenance costs

Now we obtain the expected total maintenance cost in a combined cycle, $E(C;m)$, which is the sum of OH cost, PM cost and failure repair cost,

$$E(C;m) = E(C;m;o) + E(C;m;p) + E(C;m;f). \tag{17}$$

4.2. Inventory holding cost and set-up cost

Since the production process in this case is consistent with Case 1, the calculations for inventory holding cost and set-up cost are the same as the model in Section 3, which are given in (8) and (9).

4.3. Expected profit per unit time

According to (10), we have the expected total cost during a combined cycle. Finally, based on the same gross profit in the mode of Case 1, the expected profit per unit time is obtained, as shown below

$$\begin{aligned} E(P;n,S) = & \left[(S+1)\sum_{i=1}^{k}\left(D_i R_i/n\right) - E(C;n,S)\right] \Big/ (S+1)\sum_{i=1}^{k} T_i \\ = & \Big\{(S+1)\sum_{i=1}^{k}\left(D_i R_i/n\right) \\ & -\Big\{C_o k + C_d (S+1)\sum_{i=1}^{k}\int_0^{T_i} \delta\left[1-F_1(t_1)\right]dt_1 - C_d \sum_{i=1}^{k}\int_0^{T_{i(S+1)\%k}} \delta\left[1-F_1(t_1)\right]dt_1 + C_p S k \\ & + C_f^1 (S+1)\sum_{i=1}^{k}\int_0^{T_i} \delta F_1(t_1)\,dt_1 + C_f^2 \sum_{i=1}^{k}\int_0^{TF_i} \lambda_2(t_2)\,dt_2 \\ & -(S+1)\left(\sum_{i=1}^{k} T_i\right)\left[\sum_{j=1}^{k}\left(p_j - D_j \Big/ \left(n\sum_{i=1}^{k} T_i\right)\right) T_j C_h^j\right] \Big/ 2 - (S+1)\sum_{i=1}^{k} C_s^i \Big\}\Big\} \Big/ \left[(S+1)\sum_{i=1}^{k} T_i\right], \end{aligned} \tag{18}$$

where $T_i = D_i/(np_i)$ is the production time needed for product i within a production cycle, $\sum_{i=1}^{k} T_i$ is the time of a production cycle and $\left[(S+1)\sum_{i=1}^{k} T_i\right]$ is the time of a combined cycle. Maximizing $E(P;n,S)$, we can obtain the optimal n and S.

5. Property of the models

Since this paper aims to seek both the optimal n and S that maximize $E(P;n,S)$ in equations (11) and (18), we need to know whether the optimal n and S exist or on what conditions they exist. If the optimal solution exists, one can seek it by maximizing $E(P;n,S)$ in equations (11) and (18). Moreover, the optimal solution can also be obtained according to these properties. However, the production period may last a long time period while the production cycle and maintenance cycle are just counted by a shorter time period in practice. As a result, n and S may be triple digits at most. In addition, n and S are discrete variables. Thus, the search space is usually small. Enumeration is a direct and reliable method to obtain the optimal result. Herein, the recommended solutions of optimization can be constructed based on the enumeration of the search space and the variation trend.

In this section, focusing on the existence of the two decision variables, we discuss their properties. We first study the property of optimal S for fixed n and then study the property of optimal n for fixed S. Since the properties and search algorithms for the models of two cases are similar, we just take the model of Case 1 as an example here.

5.1. Optimal S

Proposition 1. The optimal S, S^*, can be obtained in three situations:

1). If $\lim_{S\to\infty} L(P;S) > C_o\sum_{i=1}^{k} T_i - C_d\sum_{i=1}^{k} T_i\int_0^{\sum_{i=1}^{k} T_i}\delta\left[1-F_1(t_1)\right]dt_1 - C_p > L(P;2)$, there exists a finite and unique S^* which can be obtained by satisfying the inequalities

$$\begin{cases} L(P;S+1) \geq C_o\sum_{i=1}^{k} T_i - C_d\sum_{i=1}^{k} T_i\int_0^{\sum_{i=1}^{k} T_i}\delta\left[1-F_1(t_1)\right]dt_1 - C_p \\ L(P;S) < C_o\sum_{i=1}^{k} T_i - C_d\sum_{i=1}^{k} T_i\int_0^{\sum_{i=1}^{k} T_i}\delta\left[1-F_1(t_1)\right]dt_1 - C_p \end{cases}, \tag{19}$$

where $L(P;S) = C_f^2\sum_{i=1}^{k} T_i\left[S\int_{S\sum_{i=1}^{k} T_i}^{(S+1)\sum_{i=1}^{k} T_i}\lambda_2(t_2)\,dt_2 - \int_0^{S\sum_{i=1}^{k} T_i}\lambda_2(t_2)\,dt_2\right]$;

2). If $C_o\sum_{i=1}^{k} T_i - C_d\sum_{i=1}^{k} T_i\int_0^{\sum_{i=1}^{k} T_i}\delta\left[1-F_1(t_1)\right]dt_1 - C_p \leq L(P;2)$, $S^* = 1$;

3). If $C_o\sum_{i=1}^{k} T_i - C_d\sum_{i=1}^{k} T_i\int_0^{\sum_{i=1}^{k} T_i}\delta\left[1-F_1(t_1)\right]dt_1 - C_p \geq \lim_{S\to\infty} L(P;S)$, there is no need to conduct OH and the S^* does not exist.

The proof process of **Proposition 1** is presented in Appendix. According to **Proposition 1**, we can give the following Proposition straightforwardly.

Proposition 2. There exists a unique S, that maximizes $E(P;n,S)$ if $\lim_{t_2\to\infty}\lambda_2(t_2)=\infty$, for any fixed $n\geq 1$.

The proof process of **Proposition 2** is presented in Appendix. According to **Proposition 2**, if the occurrence of failure mode II follows a Weibull Distribution $f_2(t_2)=(b/a)(t_2/a)^{b-1}e^{-(t_2/a)^b}$ with $b>1$, there exists a unique S that maximizes $E(P;n,S)$ since

$$\lim_{t_2\to\infty}\lambda_2(t_2)=\lim_{t_2\to\infty}\frac{f_2(t_2)}{1-F_2(t_2)}=\lim_{t_2\to\infty}\frac{b}{a}\frac{(t_2/a)^{b-1}e^{-(t_2/a)^b}}{1-\left(1-e^{-(t_2/a)^b}\right)}=\lim_{t_2\to\infty}\frac{b}{a}\left(\frac{t_2}{a}\right)^{b-1}=\infty. \tag{20}$$

5.2. Optimal n

Proposition 3. $E(P;n,S)$ is a strictly concave function of n for any fixed $S\geq 1$, so there exists a unique optimal n, n^*, that maximizes $E(P;n,S)$. n^* can be obtained by satisfying

$$E(P;n,S)'=\left[n(S+1)\sum_{i=1}^{k}T_i\right]^{-1}\left\{M(P;n)-\left[C_o+C_pS+(S+1)\sum_{i=1}^{k}C_s^i\right]\right\}=0, \tag{21}$$

where

$$M(P;n)=M_1(P;n)+M_2(P;n)+M_3(P;n), \tag{22}$$

and

$$\begin{cases} M_1(P;n)=\left(C_f^1(S+1)\delta-C_dS\delta\right)\left[\left(\sum_{i=1}^{k}T_i\right)F_1\left(\sum_{i=1}^{k}T_i\right)-\int_0^{\sum_{i=1}^{k}T_i}F_1(t_1)\,dt_1\right],\\ M_2(P;n)=C_f^2\left[(S+1)\left(\sum_{i=1}^{k}T_i\right)\lambda_2\left((S+1)\sum_{i=1}^{k}T_i\right)-\int_0^{(S+1)\sum_{i=1}^{k}T_i}\lambda_2(t_2)\,dt_2\right],\\ M_3(P;n)=(S+1)\left[\left(\sum_{i=1}^{k}T_i\right)\sum_{j=1}^{k}\left(p_j-D_j\Big/\left(\sum_{i=1}^{k}D_i/p_i\right)\right)T_jC_h^j/2\right], \end{cases} \tag{23}$$

where $T_i=D_i/(np_i)$.

The proof process of **Proposition 3** is presented in Appendix. In practice, n^* should be an integer. Since $E(P;n,S)$ is a strictly concave function of n for any $S>0$, we just need to compare $E(P;n_1^*,S)$ and $E(P;n_2^*,S)$ to obtain the optimal n, where $n_1^*=\text{int}(n^*)$ and $n_2^*=\text{int}(n^*)+1$. Specifically, n_2^* is straightforwardly the optimal n if $n_1^*=0$ and $n_2^*=1$, since $E(P;0,S)$ does not exist. There is a small possibility that $E(P;n_1^*,S)=E(P;n_2^*,S)$, so we may potentially have two optimal values.

6. Numerical examples

In this section, two examples are presented to illustrate the models of Case 1 and Case 2, respectively. We now choose two probability distributions for two failure modes, respectively. Exponential distributions have been used for delay-time modelling for many years (Christer and Wang, 1995). In addition, previous simulations have illustrated that for a complex system with many components, the pooled delay-time approximately follows an exponential distribution (Wang, 2012). Thus, we assume that the pdf of the delay-time of defects in failure mode I is an exponential distribution with parameter φ. Besides, a Weibull distribution with scale parameter a and shape parameter b is used as the pdf of the failure time of failure mode II. Assume there is a centrifugal system where six different sizes of cast iron pipes are produced alternately. In this production process, the time unit is day and the production quantity unit is ton. Parameters used for two examples are shown in **Table 1** and **Table 2**.

Table 1. The values of parameters that are different for different products

	Parameter				
Product	D_i	p_i	C_h^i	C_s^i	R_i
1	4500	50	0.31	200	380
2	2500	50	0.33	210	410
3	4000	80	0.32	205	400
4	3600	60	0.31	200	380
5	2000	50	0.34	220	420
6	3500	50	0.32	208	400

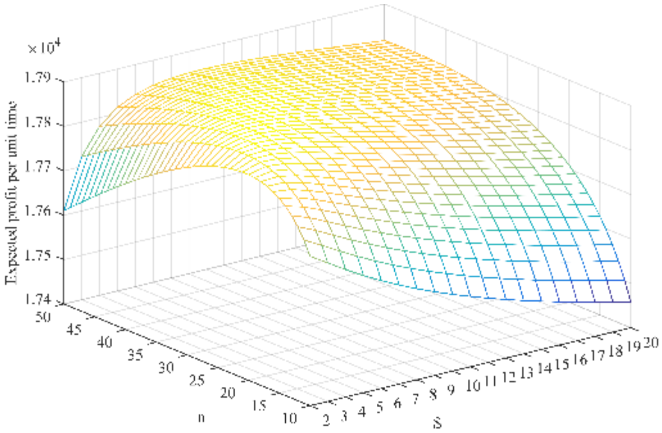
Table 2. The values of parameters that are same for different products

C_d	C_p	C_o	C_f^1	C_f^2	δ	φ	a	b
600	200	15000	1500	3000	0.225	0.042	1.03	1.05

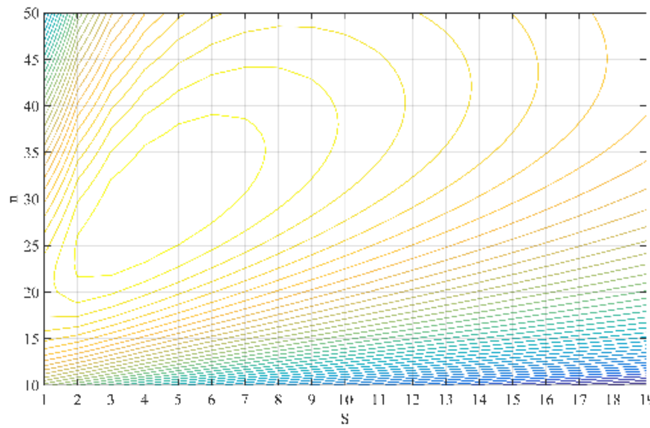
It can be seen that only 6 products are produced by the system and the least demand of these products is 2000. Herein, the maximums of n and S are 2000 and 1999 respectively, which means that the search space for this case is less than 4,000,000. In fact, the number of lots (n) will be less than 100 in practice. Anyway, the search space is not large for this case. As we discussed in Section 5, the enumeration approach that compares all the results of concerned search space is appropriate for the solution of the optimization problem of small search space. Thus, enumeration approach is used here to obtain the optimal solution.

6.1. Example 1: the example for the model of Case 1

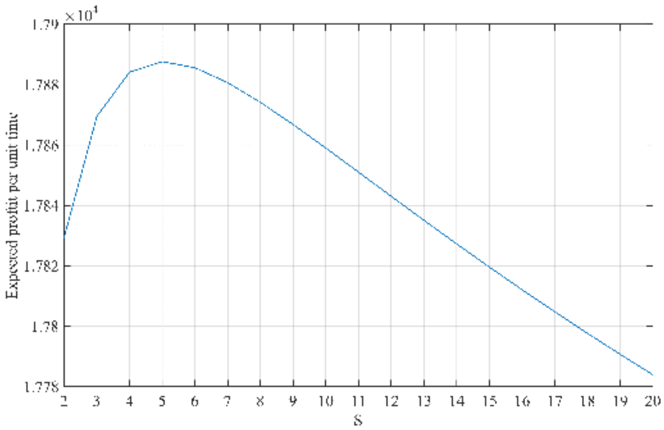
This example is presented to illustrate the model proposed for Case 1. Based on the proposed model and the values of parameters, the maximum expected profit per unit time is achieved when $n = 29$ and $S = 5$, which is 17887.66. The results of the expected profit per unit time in terms of n changing from 10 to 50 and S changing from 2 to 20 are shown in **Fig. 3**. To clearly detail the variation of the expected profit per unit time, the results are shown in four perspectives: (a) 3 dimensions, (b) contour, (c) $n=29$, and (d) $S=5$.



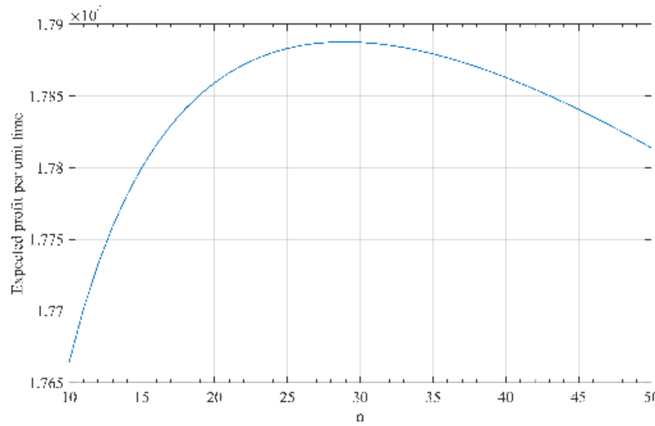
(a) 3 dimensions



(b) contour



(c) *n*=29



(d) *S*=5

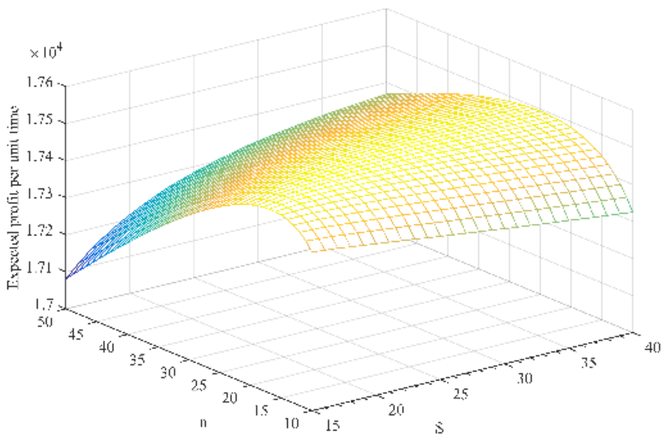
Fig. 3. The expected profit per unit time as a function of *n* and *S* by the model in Case 1

In Fig 3, it can be seen that the expected profit per unit time generally increases within small n or S and then decreases. The maximum value is 17887.66 at the point of $n=29$ and $S=5$. The results suggest the production system to produce in 30 production cycles (lots), where the optimal lot sizes for the six products are $4500/29\approx155$, $2500/29\approx86$, $4000/29\approx138$, $3600/29\approx124$, $2000/29\approx69$, $3500/29\approx121$. It is also recommended that the maintenance cycle contains $5+1=6$ production cycles, where PM is carried out at the end of each of the first 5 cycles and OH is carried out at the end of the 6th cycle. Besides, according to the model, we can know the duration of a production cycle and a maintenance cycle are about 12.41 and 74.46 days respectively.

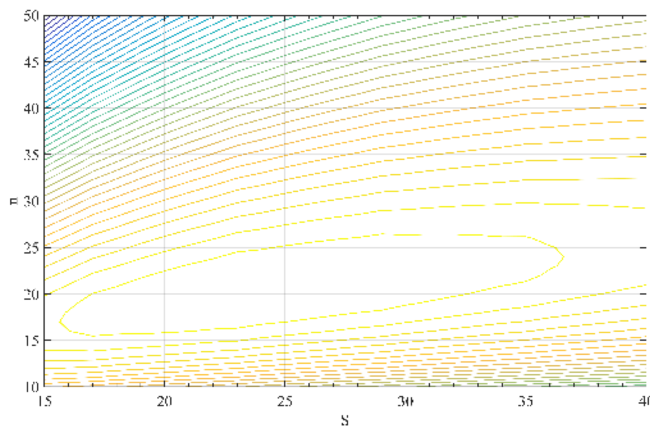
The results can be explained in a managerial logic. Basically, PM and OH become more frequent as *n* increases (shorter production cycle). Compared with abandoning maintenance, more frequent preventive actions can attenuate the risk of system failure and thus reduce the cost caused by failure, which is more effective to achieve higher profit. However, the cost of preventive actions also increases with the maintenance becoming more frequent. As a result, the profit will reduce if the maintenance is excessive. Thus, optimal n is obtained to strike the balance between the cost of preventive actions and the cost caused by failures. In contrast, PM also becomes more frequent but OH becomes less frequent as S increases, which means that the optimal S determines the optimal balance between the frequencies of PM and OH. In addition, since inspection does not prevent the occurrence of failure mode II, an appropriate frequency of OH is needed to prevent the system from failures of mode II. However, since the cost of an OH is much higher than the cost of a PM, too intensive OH is also an excessive maintenance which brings extra expenditure and reduces the profit. Herein, both n and S should be optimized in practice.

6.2. Example 2: the example for the model of Case 2

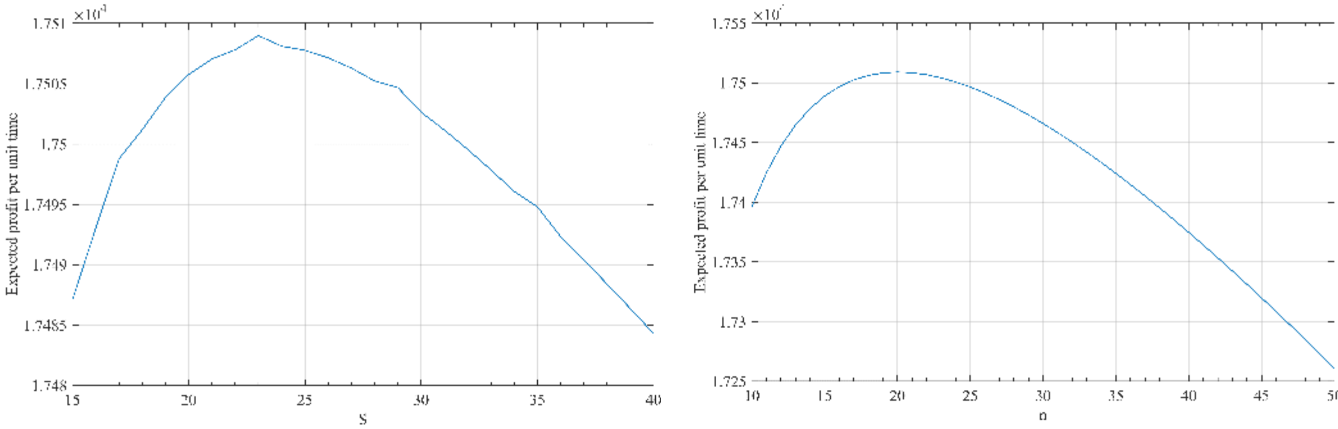
This example is presented to illustrate the model proposed for Case 2. The parameters here are set as the same as in Example 1. Based on the proposed model, the maximum expected profit per unit time is achieved when $n = 20$ and $S = 23$, which is 17508.98. The results of the expected profit per unit time in terms of n changing from 10 to 50 and S changing from 15 to 40 are shown in **Fig. 4**. To clearly detail the variation of the expected profit per unit time, the results are shown in four perspectives: (a) 3 dimensions, (b) contour, (c) $n = 20$, (d) $S = 23$.



(a) 3 dimensions



(b) contour



(c) n=20 (d) S=23

Fig. 4. The expected profit per unit time as a function of n and S by the model in Case 2

It can be seen that the changing trend of the expected profit per unit time by n and S in example 2 is the same as in example 1. The maximum result is 17508.46 at the point of $n = 20$ and $S = 23$. This means the production system is suggested to produce in 20 production cycles(lots), where the optimal lot sizes for the six products are $4500/20=225$, $2500/20=125$, $4000/20=200$, $3600/20=180$, $2000/20=100$, $3500/20=175$. The maintenance cycle is suggested to include $23+1=24$ set-ups, where PM is carried out at each of the first 23 set-up points and OH is carried out at the 24th set-up point. Besides, according to the model, we can know the duration of a production cycle is 18 while the duration of a combined cycle is $18\times 24=432$. As the maintenance cycle is not constant in the model of Case 2, we calculate the mean duration of a maintenance cycle as a reference, which is $432/6=72$.

Comparing the results of these two examples, we can see that the maximum expected profit per unit time calculated by the model of Case 1 is larger than that of Case 2. This is because that maintaining the system at each set-up point is an excessive maintenance policy, which largely increases the maintenance cost. Herein, in this case it is better to maintain the system at the end of the production cycle. Besides, we can also find that the length of a production cycle in example 2 is longer than that that in example 1 (18 vs 12.41). The reason behind is that in case of excessive maintenance, the interval time between set-ups is prolonged in the optimal policy in example 2. In contrast, since PM does not affect the failure process of mode II, the difference in the intervals of OH in optimal solution for two maintenance policies is small. As a result, the duration of maintenance cycle is generally the same for two examples (72 vs 74.46).

In addition, it should be noted that OH will not be carried out if $S = +\infty$. In this case, the maintenance policy can only prevent the failure mode I and thus will largely decrease the profit, which is illustrated by the trend of the results above. Although lot sizing and maintenance policy can be optimized only considering the soft failure (failure mode I), it cannot obtain the maximal profit if the system subjects to both the soft failure and the hard failure. In fact, only under some extreme conditions the optimization considering only soft failures can reach the same result as optimization considering both soft failures and hard failures, which is discussed in the following sensitivity analyses. Herein, compared with the previous model where only the failure mode I was considered (Liu et al., 2015), the proposed model in this study is significantly improved to be applied widely.

6.3. Sensitivity analyses

Here, sensitivity analyses are performed to investigate the sensitivities of the optimal schedule and the corresponding expected profit per unit time with respect to the variations of the parameters used. For sensitivity analyses, we use the original values of parameters in **Table 1** and **Table 2** as base-case values to adjust these parameters, where high value is 1.5 times the base-case value and low value is 0.5 times the base-case value. Then, we obtain the optimal lot sizing and maintenance policy as well as the corresponding expected profit per unit time, based on the adjusted parameters one at a time. To present the variance, we calculate the difference between the results of high value and low value. The optimal lot sizing and maintenance policy as well as the corresponding expected profit per unit time obtained based on base-case value are also listed as a reference. To facilitate the comparison between different parameters, we further present the change ratio which is defined as the difference of the expected profit per unit time dividing by the difference between high value and low value of parameter.

Since a production cycle is a complete run of all products, the parameters for different products in **Table 1** are synchronously varied for all products. Because there are 6 different products for the same parameter, the change ratio of the parameters in **Table 1** is calculated as the difference of the expected profit per unit time dividing by the mean difference between high value and low value of parameter. The results of sensitivity analysis for Example 1 and Example 2 are shown in **Table 3** and **Table 4,** respectively.

Table 3. Sensitivity analysis for Example 1

	Optimal schedule (n, S), the expected profit per unit time				
Parameters	Low value	Base-case value	High value	Difference	Change ratio
D_i	(14, 5), 17887.44	(29, 5), 17887.66	(43, 5), 17887.65	(29, 0), 0.21	0.00006
p_i	(47, 4), 6907.07	(29, 5), 17887.66	(23, 6), 28877.57	(-24, 2), 21970.50	387.71471
C_h^i	(23, 4), 17938.97	(29, 5), 17887.66	(34, 6), 17845.71	(11, 2), -93.26	-289.92746
C_s^i	(38, 7), 17944.75	(29, 5), 17887.66	(24, 4), 17842.30	(-14, -3), -102.45	-0.494523
R_i	(29, 5), 6855.71	(29, 5), 17887.66	(29, 5), 28919.60	(0, 0), 22063.89	55.39052
C_d	(32, 6), 17932.42	(29, 5), 17887.66	(26, 4), 17844.98	(-6, -2), -84.74	-0.14123
C_p	(29, 5), 17894.37	(29, 5), 17887.66	(28, 5), 17880.96	(-1, 0), -13.41	-0.00671
C_o	(29, 2), 18030.77	(29, 5), 17887.66	(29, 8), 17807.02	(0, 6), -223.74	-0.01492
C_f^1	(25, 4), 17927.34	(29, 5), 17887.66	(33, 6), 17852.30	(8, 2), -75.04	-0.05003
C_f^2	(30, 11), 19718.64	(29, 5), 17887.66	(28, 3), 16101.93	(-2, -8), -3616.71	-1.20557

δ	(28, 5), 17969.26	(29, 5), 17887.66	(30, 5), 17806.97	(2, 0), -16.23	-72.13333
φ	(25, 4), 17911.05	(29, 5), 17887.66	(30, 5), 17868.49	(5, 1), -42.56	-1013.33333
a	(28, 2), 14098.82	(29, 5), 17887.66	(29, 8), 19150.65	(1, 6), 5051.84	-4904.69903
b	(31, +Inf), 21676.97	(29, 5), 17887.66	(160,1), 11237.40	(129, -Inf), -10439.57	-9942.44762

Table 4. Sensitivity analysis for Example 2

	Optimal schedule (n, S), the expected profit per unit time				
Parameters	Low value	Base-case value	High value	Difference	Change ratio
D_i	(10, 23), 17508.98	(20, 23), 17508.98	(30, 23), 17508.98	(20, 0), 0	0
p_i	(30, 17), 6552.86	(20, 23), 17508.98	(16, 29), 28481.79	(-14, 12), 21928.93	386.98111
C_h^i	(15, 17), 17584.80	(20, 23), 17508.98	(25, 29), 17449.94	(10, 12), -134.86	419.25389
C_s^i	(24, 29), 17546.44	(20, 23), 17508.98	(19, 23), 17475.49	(-5, -6), -70.96	0.34252
R_i	(20, 23), 6477.04	(20, 23), 17508.98	(20, 23), 28540.93	(0, 0), 22063.89	55.39052
C_d	(20, 23), 17569.08	(20, 23), 17508.98	(20, 23), 17448.89	(0, 0), -120.19	-0.20032
C_p	(24, 29), 17543.68	(20, 23), 17508.98	(19, 23), 17477.94	(-5, -6), -65.73	-0.32865
C_o	(20, 11), 17652.05	(20, 23), 17508.98	(20, 35), 17425.35	(0, 24), -226.70	-0.01511
C_f^1	(20, 23), 17519.87	(20, 23), 17508.98	(20, 23), 17498.10	(0, 0), -21.77	-0.01451
C_f^2	(20, 47), 19508.89	(20, 23), 17508.98	(21, 17), 15557.88	(1, -30), -3951.01	-1.31700
δ	(20, 23), 17579.97	(20, 23), 17508.98	(20, 23), 17438.00	(0, 0), -141.97	-630.97778
φ	(20, 23), 17515.59	(20, 23), 17508.98	(20, 23), 17502.69	(0, 0), -12.89	-306.90476
a	(20, 11), 13370.26	(20, 23), 17508.98	(20, 35), 18888.56	(0, 24), -5518.30	-5357.5728
b	(20, +Inf), 21647.71	(20, 23), 17508.98	(56, 1), 2220.66	(36, -Inf), -19427.05	-18501.96

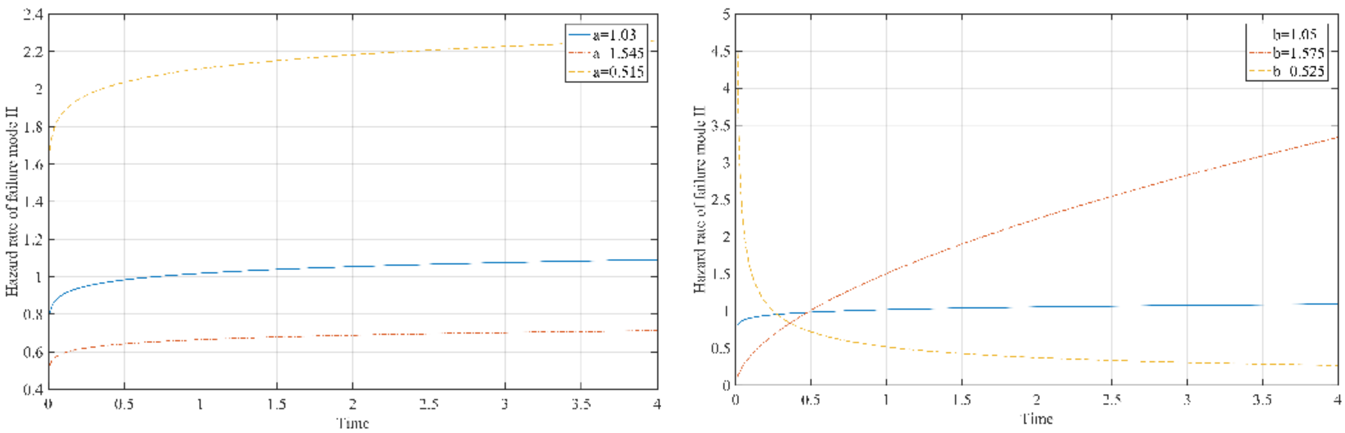
Table 3 and **Table 4** show that the optimal result is sensitive with respect to the changes in all the parameters regarding products. Particularly, when the demands of these products increase (D_i increase), the expected profit per unit time can remain unchanged if the number of lots, n, also increases. This is reflected as n increases but the expected profit per unit time is almost constant with the growth in demands. On the contrary, the variation of the gross profit per unit product (R_i) cannot affect the optimal n and S, but can directly improve the expected profit per unit time. Increasing the production rate (increasing p_i) can also improve the expected per unit time but the production cycle needs to be prolonged (n decreases) simultaneously so that the production is interrupted less frequently. Increasing both inventory holding cost (C_h^i) and set-up cost (C_s^i) will reduce the expected profit per unit time. When the inventory holding cost rises, the production cycle needs to be shortened (n increases) to reduce the inventory holding in order to improve the profit. Oppositely, when the set-up cost rises, the production cycle needs to be prolonged (n decreases) to reduce the number of set-ups in order to improve the profit.

Since the proposed model aims to study how to integrate the maintenance policy with lot sizing, the focus of the sensitivity analyses is the parameters regarding maintenance, which are shown in **Table 2**. From **Table 3** and **Table 4,** we know that the variations of all the cost parameters (C_d, C_p, C_o, C_f^1, C_f^2) can change the maximum of the expected profit per unit time. It is easy to see that the expected profit per unit time changes most significantly with respect to the variation of C_f^2 among the parameters related to cost. In particular, one cost unit increased in repairing a failure of mode II will decrease 1.20557 and 1.31700 expected profit per time unit under the optimal lot sizing and maintenance policy obtained by models of Case 1 and Case 2, respectively, while this figure is less than 0.5 for other cost parameters. This implies us that reducing the cost of repairing failures is the effective way to increase profit.

The optimal lot sizing and maintenance policy is also sensitive with respect to the changes in these cost parameters. A larger C_d and C_p will reduce the optimal *n* and *S*, which means that longer production cycle and shorter maintenance cycle or combined cycle are recommended so that the frequency of PM is reduced if the costs of repairing a defective component and an inspection are higher. Oppositely, a larger C_o will increase optimal *n* and *S*, which means that shorter production cycle and longer maintenance cycle or combined cycle are recommended. In this case, OH is less frequent (larger *S*) to avoid high cost of overhaul, whereas the frequency of PM is increased (larger *n*) to maintain the production system. For failure cost parameters, the optimal *n* and *S* increase with C_f^1 but decrease with C_f^2. This paradox can be due to two reasons. On the one hand, a shorter production cycle (larger *n*) and a longer maintenance cycle or combined cycle (larger *S*) will increase the frequency of PM, which can avoid more failures of mode I if the cost of repairing a failure of this mode is larger (larger C_f^1). On the other hand, a shorter maintenance cycle or combined cycle (smaller *S*) will increase the frequency of OH, which can avoid more failures of mode II if the cost of repairing a failure of this mode is higher (larger C_f^1). However, the results also prove that if the frequency of one preventive action is increased, the frequency of the other preventive action should be appropriately reduced in case of excessive maintenance.

The results of the proposed models are also sensitive with respect to the changes in the values of the parameters about the failure modes (δ, φ, a, b). It is obvious that the increases in the values of these parameters will reduce the expected profit per unit time. In fact, these increases lead to more possible failures so that a more intensive maintenance policy is needed to prevent the production system from these failures. In detail, the increase of δ indicates the increase in the occurrence rate for the arrival of defects, and the increase of φ indicates the decrease in the delay time for the arrival of failure in mode I. In this case, PM should be more frequent in case of failure mode I, which means a shorter production cycle (larger *n*) is recommended. Similarly, the decrease of *a* from 1.545 to 0.515 indicates that the hazard rate of failure in mode II is larger, as shown in **Fig. 5 (a),** and thus OH should be conducted more frequently (smaller *S*).

However, it is noted that the situation for b is different, as shown in **Fig. 5 (b)**. If $b<1$, the failure rate decreases with operating time, which means that there is no need to carry out OH and S does not exist. This is consistent with **Proposition 2** in Section 5.1. Specifically, the failure rate is extremely large in the beginning if b is too small so that no matter how intensive OH is carried out, the failures of mode II still cannot be prevented. As a result, the optimal maintenance policy is to abandon OH in order to save the maintenance cost, as shown in two tables (S=+Inf means that the maintenance cycle is infinite). Oppositely, If $b>1$, the failure rate increases with operating time and the increasing rate is faster when b is larger. Thus, the OH should be carried out more intensive (larger n and smaller S) to renew the system in order to limit the increasing of the failure rate. It can be seen that only under the extreme conditions such as the concentration of hard failure risk at the beginning of the study period, it is unnecessary to consider both two failure modes. In this case, the effect of the proposed model is the same as the effect of previous models; otherwise, our model is better to handle the problem of the joint optimization of the lot sizing and maintenance policy.



(a) Hazard rate of failure mode II with different a

(b) Hazard rate of failure mode II with different b

Fig. 5. Hazard rate of failure mode II with different a and b

7. Conclusion

In this paper, we studied the problem of joint optimization of lot sizing and maintenance policy for a multi-product production system subject to two failure modes. With some proposed propositions, we proved that the optimal solution exists under some certain conditions that are easy to be satisfied in practice. Numerical examples and sensitivity analyses are presented to illustrate the applications. The results of the examples show that lot sizing and maintenance policy need to be optimized in order to achieve the maximum of the profit. Both the shortage and the excess of maintenance will lead to the reduction of the expected profit per unit time. From the sensitivity analysis, it is found that reducing the cost caused by failures and improving the system reliability are effective ways to increase the expected profit per unit time.

Acknowledgement

This work was supported in part by the National Natural Science Foundation of China under Grant No. 71671016.

References

Barata, J., Soares, C. G., Marseguerra, M., & Zio, E. (2002). Simulation modelling of repairable multi-component deteriorating systems for ‘on condition’ maintenance optimization. *Reliability Engineering & System Safety, 76*(3), 255-264.

Ben-Daya, M., & Noman, S. A. (2006). Lot sizing, preventive maintenance, and warranty decisions for imperfect production systems. *Journal of Quality in Maintenance Engineering, 12*(1), 68-80.

Ben-Daya, M., Darwish, M., & Ertogral, K. (2008). The joint economic lot sizing problem: review and extensions. *European Journal Operational Research, 185*(2), 726–42.

Cha, J. H., Finkelstein, M., & Levitin, G. (2017). On preventive maintenance of systems with lifetimes dependent on a random shock process. *Reliability Engineering & System Safety, 168*, 90-97.

Cha, J. H., Finkelstein, M., & Levitin, G. (2018). Bivariate preventive maintenance of systems with lifetimes dependent on a random shock process. *European Journal of Operational Research, 266*(1), 122-134.

Cheng, G. Q., Zhou, B. H., & Li, L. (2018). Integrated production, quality control and condition-based maintenance for imperfect production systems. *Reliability Engineering & System Safety, 175*, 251-264.

Christer, A. H., & Wang, W. (1995). A delay-time-based maintenance model of a multi-component system. *IMA Journal of Management Mathematics, 6*(2), 205-222.

Garbatov, Y., & Soares, C. G. (2001). Cost and reliability based strategies for fatigue maintenance planning of floating structures. *Reliability Engineering & System Safety, 73*(*3*), 293-301.

Kilic, O. A., & Tunc, H. (2019). Heuristics for the stochastic economic lot sizing problem with remanufacturing under backordering costs. *European Journal of Operational Research,276*(3), 880-892.

Lamas, A., & Chevalier, P. (2018). Joint dynamic pricing and lot-sizing under competition. *European Journal of Operational Research,266*(3), 864-876.

Levitin, G., & Lisnianski, A. (2000). Optimization of imperfect preventive maintenance for multi-state systems. *Reliability Engineering & System Safety, 67*(2), 193-203.

Li, P., Wang, W., & Peng, R. (2016). Age-based replacement policy with consideration of production wait time. *IEEE Transactions on Reliability, 65*(1), 235-247.

Liao, G. (2012).Optimal economic production quantity policy for randomly failing process with minimal repair, back order and preventive maintenance. *International Journal of Systems Science, 44*(9), 1602-1612.

Liu, B., Lin, J., Zhang, L., & Kumar, U. (2019). A dynamic prescriptive maintenance model considering system aging and degradation. *IEEE Access, 7*, 94931 - 94943.

Liu, X., Wang, W., & Peng, R. (2015). An integrated production, inventory and preventive maintenance model for a multi-product production system.*Reliability Engineering and System Safety, 137*, 76-86.

Lu, Z., Zhang, Y., & Han, X. (2013). Integrating run-based preventive maintenance into the capacitated lot sizing problem with reliability constraint. *International Journal of Production Research, 51*(5), 1379-1391.

Mahmoudi, M., Elwany, A., Shahanaghi, K., & Gholamian,M. R. (2017). A delay time model with multiple defect types and multiple inspection methods. *IEEE Transactions on Reliability, 66*(4), 1073-1084.

Nguyen, H. S. H., Do, P., Vu, H., &I, B. (2019). Dynamic maintenance grouping and routing for geographically dispersed production systems. *Reliability Engineering & System Safety, 185*, 392-404.

Ou, J., & Feng, J. (2019). Production lot-sizing with dynamic capacity adjustment. *European Journal of Operational Research,227*(1), 261-269.

Peng, H., Q. Feng, & D. W. Coit. (2010). Reliability and maintenance modeling for systems subject to multiple dependent competing failure processes. *IIE Transactions, 43*(1), 12-22.

Peng, R., Liu, B., Zhai, Q., & Wang, W. (2019). Optimal maintenance strategy for systems with two failure modes. *Reliability Engineering & System Safety, 188*, 624-632.

Rivera-Gómez, H., Gharbi, A., Kenné, J. P., Montaño-Arango, O., & Corona-Armenta, J. R. (2020). Joint optimization of production and maintenance strategies considering a dynamic sampling strategy for a deteriorating system. *Computers & Industrial Engineering, 140*, 106273.

Schreiber, M., Vernickel, K., Richter, C., & Reinhart, G. (2019). Integrated production and maintenance planning in cyber-physical production systems. *Procedia CIRP, 79*,534-539.

Taş, D., Gendreau, M., Jabali, O., & Jans, R. (2019). A capacitated lot sizing problem with stochastic setup times and overtime. *European Journal of Operational Research,273*(1), 146-159.

Wang, H., Wang, W., & Peng, R. (2017). A two-phase inspection model for a single component system with three-stage degradation. *Reliability Engineering & System Safety, 158*, 31-40.

Wang, W. (2012). An overview of the recent advances in delay-time-based maintenance modelling. *Reliability Engineering & System Safety, 106*(5), 165-178.

Wang,W., & Wu, S. (2014). An inspection model subject to small stoppages and hard failures. *Quality Technology & Quantitative Management, 11*(3), 255-264.

Wu, S., & Clements-Croome, D. (2005). Optimal maintenance policies under different operational schedules. *IEEE Transactions on Reliability,54*(2), 338-346.

Wu, S., & Zuo, M. J. (2010). Linear and Nonlinear Preventive Maintenance Models. *IEEE Transactions on Reliability, 59*(1), 242-249.

Xiao, H., & Peng, R. (2014). Optimal allocation and maintenance of multi-state elements in series–parallel systems with common bus performance sharing. *Computers & Industrial Engineering, 72*, 143-151.

Yang, L., Ye, Z. S., Lee, C. G., Yang, S. F., & Peng, R. (2019). A two-phase preventive maintenance policy considering imperfect repair and postponed replacement. *Reliability Engineering & System Safety, 274*(3), 966-977.

Yang, L., Zhao, Y., Peng, R., & Ma, X. (2018). Hybrid preventive maintenance of competing failures under random environment. *Reliability Engineering & System Safety*, *174*, 130-140.

Ye, Z., Revie, M., & Walls, L. (2014). A load sharing system reliability model with managed component degradation. *IEEE Transactions on Reliability*, *63*(3), 721-730.

Zhang, M., Ye, Z., & Xie, M. (2014). A condition-based maintenance strategy for heterogeneous populations. *Computers & Industrial Engineering*, *77*, 103-114.

Zhao, F., Wang, W., & Peng, R. (2015). Delay-time-based preventive maintenance modelling for a production plant: a case study in a steel mill. *Journal of the Operational Research Society*, *66*(12), 2015-2024.

Zhou, X., & Yu, M. (2020). Semi-dynamic maintenance scheduling for multi-station series systems in multi-specification and small-batch production. *Reliability Engineering & System Safety*, *195*, 106753.

Zhou, Y., Guo, Y., Lin, T. R., &Ma, L. (2018). Maintenance optimisation of a series production system with intermediate buffers using a multi-agent FMDP. *Reliability Engineering & System Safety*, *180*, 39-48.

Appendix

A.1. Proof of Proposition 1

To find the optimal S^* that maximises $E(P;n,S)$, we form the inequalities

$$\begin{cases} E(P;n,S) \ge E(P;n,S+1) \\ E(P;n,S-1) < E(P;n,S) \end{cases}, \tag{A1}$$

From equation (A1), we can obtain

$$\begin{cases} \left[(S+1)\sum_{i=1}^{k}(D_i R_i/n) - E(C;n,S)\right] \Big/ (S+1)\sum_{i=1}^{k} T_i \\ \ge \left[(S+1+1)\sum_{i=1}^{k}(D_i R_i/n) - E(C;n,S+1)\right] \Big/ (S+1+1)\sum_{i=1}^{k} T_i , \\ \left[(S-1+1)\sum_{i=1}^{k}(D_i R_i/n) - E(C;n,S-1)\right] \Big/ (S-1+1)\sum_{i=1}^{k} T_i \\ \le \left[(S+1)\sum_{i=1}^{k}(D_i R_i/n) - E(C;n,S)\right] \Big/ (S+1)\sum_{i=1}^{k} T_i , \end{cases} \tag{A2}$$

that is,

$$\begin{cases} E(C;n,S) \Big/ (S+1)\sum_{i=1}^{k} T_i \le E(C;n,S+1) \Big/ (S+2)\sum_{i=1}^{k} T_i , \\ E(C;n,S-1) \Big/ S\sum_{i=1}^{k} T_i \ge E(C;n,S) \Big/ (S+1)\sum_{i=1}^{k} T_i . \end{cases} \tag{A3}$$

The first inequality can be reformed as

$$E(C;n,S)(S+2)\sum_{i=1}^{k} T_i - E(C;n,S+1)(S+1)\sum_{i=1}^{k} T_i \le 0. \tag{A4}$$

Specifically, we have

$$
\begin{aligned}
&E(C;n,S)\left(S+2\right)\sum\nolimits_{i=1}^{k}T_i - E(C;n,S+1)\left(S+1\right)\sum\nolimits_{i=1}^{k}T_i \\
&=\left\{\left\{C_o + C_d S\int_0^{\sum_{i=1}^{k}T_i}\delta\left[1-F_1\left(t_1\right)\right]dt_1 + C_p S + C_f^1\left(S+1\right)\int_0^{\sum_{i=1}^{k}T_i}\delta F_1\left(t_1\right)dt_1 + C_f^2\int_0^{(S+1)\sum_{i=1}^{k}T_i}\lambda_2\left(t_2\right)dt_2\right\}\right. \\
&\left.+\left\{\left(S+1\right)\left(\sum\nolimits_{i=1}^{k}T_i\right)\left[\sum\nolimits_{j=1}^{k}\left(p_j - D_j\Big/\left(n\sum\nolimits_{i=1}^{k}T_i\right)\right)T_j C_h^j\right]\Big/2\right\}+\left\{\left(S+1\right)\sum\nolimits_{i=1}^{k}C_s^i\right\}\right\}\left(S+2\right)\sum\nolimits_{i=1}^{k}T_i \\
&-\left\{\left\{C_o + C_d\left(S+1\right)\int_0^{\sum_{i=1}^{k}T_i}\delta\left[1-F_1\left(t_1\right)\right]dt_1 + C_p\left(S+1\right) + C_f^1\left(S+2\right)\int_0^{\sum_{i=1}^{k}T_i}\delta F_1\left(t_1\right)dt_1 + C_f^2\int_0^{(S+2)\sum_{i=1}^{k}T_i}\lambda_2\left(t_2\right)dt_2\right\}\right. \\
&\left.+\left\{\left(S+2\right)\left(\sum\nolimits_{i=1}^{k}T_i\right)\left[\sum\nolimits_{j=1}^{k}\left(p_j - D_j\Big/\left(n\sum\nolimits_{i=1}^{k}T_i\right)\right)T_j C_h^j\right]\Big/2\right\}+\left\{\left(S+2\right)\sum\nolimits_{i=1}^{k}C_s^i\right\}\right\}\left(S+1\right)\sum\nolimits_{i=1}^{k}T_i \\
&= C_o\sum\nolimits_{i=1}^{k}T_i - C_d\sum\nolimits_{i=1}^{k}T_i\int_0^{\sum_{i=1}^{k}T_i}\delta\left[1-F_1\left(t_1\right)\right]dt_1 - C_p \\
&+C_f^2\left(S+2\right)\sum\nolimits_{i=1}^{k}T_i\int_0^{(S+1)\sum_{i=1}^{k}T_i}\lambda_2\left(t_2\right)dt_2 - C_f^2\left(S+1\right)\sum\nolimits_{i=1}^{k}T_i\int_0^{(S+2)\sum_{i=1}^{k}T_i}\lambda_2\left(t_2\right)dt_2 \\
&= C_o\sum\nolimits_{i=1}^{k}T_i - C_d\sum\nolimits_{i=1}^{k}T_i\int_0^{\sum_{i=1}^{k}T_i}\delta\left[1-F_1\left(t_1\right)\right]dt_1 - C_p \\
&+C_f^2\sum\nolimits_{i=1}^{k}T_i\left[\int_0^{(S+1)\sum_{i=1}^{k}T_i}\lambda_2\left(t_2\right)dt_2 - \left(S+1\right)\int_{(S+1)\sum_{i=1}^{k}T_i}^{(S+2)\sum_{i=1}^{k}T_i}\lambda_2\left(t_2\right)dt_2\right].
\end{aligned}
\tag{A5}
$$

According to equation (A4), equation (A5) implies

$$
\begin{aligned}
&C_f^2\sum\nolimits_{i=1}^{k}T_i\left[\left(S+1\right)\int_{(S+1)\sum_{i=1}^{k}T_i}^{(S+2)\sum_{i=1}^{k}T_i}H_2\left(t_2\right)dt_2 - \int_0^{(S+1)\sum_{i=1}^{k}T_i}\lambda_2\left(t_2\right)dt_2\right] \\
&\geq C_o\sum\nolimits_{i=1}^{k}T_i - C_d\sum\nolimits_{i=1}^{k}T_i\int_0^{\sum_{i=1}^{k}T_i}\delta\left[1-F_1\left(t_1\right)\right]dt_1 - C_p.
\end{aligned}
\tag{A6}
$$

Similarly, from the second inequality in equation (A3), we have

$$
\begin{aligned}
&C_f^2\sum\nolimits_{i=1}^{k}T_i\left[S\int_{S\sum_{i=1}^{k}T_i}^{(S+1)\sum_{i=1}^{k}T_i}H_2\left(t_2\right)dt_2 - \int_0^{S\sum_{i=1}^{k}T_i}\lambda_2\left(t_2\right)dt_2\right] \\
&\leq C_o\sum\nolimits_{i=1}^{k}T_i - C_d\sum\nolimits_{i=1}^{k}T_i\int_0^{\sum_{i=1}^{k}T_i}\delta\left[1-F_1\left(t_1\right)\right]dt_1 - C_p.
\end{aligned}
\tag{A7}
$$

Thus, let $L(P;n,S) = C_f^2\sum\nolimits_{i=1}^{k}T_i\left[S\int_{S\sum_{i=1}^{k}T_i}^{(S+1)\sum_{i=1}^{k}T_i}\lambda_2\left(t_2\right)dt_2 - \int_0^{S\sum_{i=1}^{k}T_i}\lambda_2\left(t_2\right)dt_2\right]$, $S = 0,1,2,...$ we have

$$
\begin{cases}
L(P;n,S+1) \geq C_o\sum\nolimits_{i=1}^{k}T_i - C_d\sum\nolimits_{i=1}^{k}T_i\int_0^{\sum_{i=1}^{k}T_i}\delta\left[1-F_1\left(t_1\right)\right]dt_1 - C_p, \\
L(P;n,S) \leq C_o\sum\nolimits_{i=1}^{k}T_i - C_d\sum\nolimits_{i=1}^{k}T_i\int_0^{\sum_{i=1}^{k}T_i}\delta\left[1-F_1\left(t_1\right)\right]dt_1 - C_p.
\end{cases}
\tag{A8}
$$

Then since $\lambda_2\left(t_2\right)$ is usually a monotonous increasing function in practice, we evidently have

$$
\begin{aligned}
&L(P;S+1)-L(P;S)\\
&=C_f^2\sum_{i=1}^{k}T_i\left[(S+1)\int_{(S+1)\sum_{i=1}^{k}T_i}^{(S+2)\sum_{i=1}^{k}T_i}\lambda_2(t_2)dt_2-\int_0^{(S+1)\sum_{i=1}^{k}T_i}\lambda_2(t_2)dt_2\right]\\
&\quad -C_f^2\sum_{i=1}^{k}T_i\left[S\int_{S\sum_{i=1}^{k}T_i}^{(S+1)\sum_{i=1}^{k}T_i}\lambda_2(t_2)dt_2-\int_0^{S\sum_{i=1}^{k}T_i}\lambda_2(t_2)dt_2\right]\\
&=C_f^2\sum_{i=1}^{k}T_i\left[S\int_{(S+1)\sum_{i=1}^{k}T_i}^{(S+2)\sum_{i=1}^{k}T_i}\lambda_2(t_2)dt_2-S\int_{S\sum_{i=1}^{k}T_i}^{(S+1)\sum_{i=1}^{k}T_i}\lambda_2(t_2)dt_2\right.\\
&\quad\left.+\int_{(S+1)\sum_{i=1}^{k}T_i}^{(S+2)\sum_{i=1}^{k}T_i}\lambda_2(t_2)dt_2-\int_0^{(S+1)\sum_{i=1}^{k}T_i}\lambda_2(t_2)dt_2+\int_0^{S\sum_{i=1}^{k}T_i}\lambda_2(t_2)dt_2\right]\\
&=C_f^2\sum_{i=1}^{k}T_i\left[S\left(\int_{(S+1)\sum_{i=1}^{k}T_i}^{(S+2)\sum_{i=1}^{k}T_i}\lambda_2(t_2)dt_2-\int_{S\sum_{i=1}^{k}T_i}^{(S+1)\sum_{i=1}^{k}T_i}\lambda_2(t_2)dt_2\right)\right.\\
&\quad\left.+\int_{(S+1)\sum_{i=1}^{k}T_i}^{(S+2)\sum_{i=1}^{k}T_i}\lambda_2(t_2)dt_2-\int_{S\sum_{i=1}^{k}T_i}^{(S+1)\sum_{i=1}^{k}T_i}\lambda_2(t_2)dt_2\right]\\
&=C_f^2\left(\sum_{i=1}^{k}T_i\right)(S+1)\left[\int_{(S+1)\sum_{i=1}^{k}T_i}^{(S+2)\sum_{i=1}^{k}T_i}\lambda_2(t_2)dt_2-\int_{S\sum_{i=1}^{k}T_i}^{(S+1)\sum_{i=1}^{k}T_i}\lambda_2(t_2)dt_2\right]\\
&>0,
\end{aligned}
\tag{A9}
$$

and

$$
\begin{aligned}
&\lim_{S\to\infty}L(P;S)\\
&=\lim_{S\to\infty}C_f^2\sum_{i=1}^{k}T_i\left[S\int_{S\sum_{i=1}^{k}T_i}^{(S+1)\sum_{i=1}^{k}T_i}\lambda_2(t_2)dt_2-\int_0^{S\sum_{i=1}^{k}T_i}\lambda_2(t_2)dt_2\right]\\
&=\lim_{S\to\infty}C_f^2\left(\sum_{i=1}^{k}T_i\right)\sum_{j=1}^{S}\left[\int_{S\sum_{i=1}^{k}T_i}^{(S+1)\sum_{i=1}^{k}T_i}\lambda_2(t_2)dt_2-\int_{(j-1)\sum_{i=1}^{k}T_i}^{j\sum_{i=1}^{k}T_i}\lambda_2(t_2)dt_2\right].
\end{aligned}
\tag{A10}
$$

Thus, we can obtain the optimal S in the following three situations.

The first one is the situation where

$$\lim_{S\to\infty}L(P;S)>C_o\sum_{i=1}^{k}T_i-C_d\sum_{i=1}^{k}T_i\int_0^{\sum_{i=1}^{k}T_i}\delta\left[1-F_1(t_1)\right]dt_1-C_p>L(P;2). \tag{A11}$$

In this situation, there exists a finite and unique S^* which can be obtained by satisfying equation (A8) which is also equation (19).

The second one is the situation where

$$C_o\sum_{i=1}^{k}T_i-C_d\sum_{i=1}^{k}T_i\int_0^{\sum_{i=1}^{k}T_i}\delta\left[1-F_1(t_1)\right]dt_1-C_p\le L(P;2). \tag{A12}$$

In this situation, it is easy to know that $E(P;n,S)$ decreases with S, so we have $S^*=1$.

The third one is the situation where

$$C_o\sum_{i=1}^{k}T_i-C_d\sum_{i=1}^{k}T_i\int_0^{\sum_{i=1}^{k}T_i}\delta\left[1-F_1(t_1)\right]dt_1-C_p\ge\lim_{S\to\infty}L(P;S). \tag{A13}$$

In this situation, $E(P;n,S)$ continuously increases with S so $S^* \to \infty$, which means that there is no need to conduct overhaul and we do not have the S^*.

This establishes **Proposition 1**.

A.2. Proof of Proposition 2

If $\lim_{t_2 \to \infty} \lambda_2(t_2) = \infty$, according to equation (A10) we have $\lim_{S \to \infty} L(P;S) = \infty$. Then the third situation in **Proposition 1** cannot be satisfied. Thus, there exists a unique S that maximizes $E(P;n,S)$.

A.3. Proof of Proposition 3

If we treat n as a continuous variable, then $E(P;n,S)$ is a continuous function. To find an n^* which maximizes $E(P;n,S)$, we differentiate $E(P;n,S)$ with respect to n as

$$\begin{aligned} E(P;n,S)' &= \left\{ \left[(S+1)\sum_{i=1}^{k}(D_i R_i/n) - E(C;n,S) \right] \Big/ (S+1)\sum_{i=1}^{k} T_i \right\}' \\ &= \left[(S+1)\sum_{i=1}^{k} T_i \right]^{-2} \left\{ \left[(S+1)\sum_{i=1}^{k}(D_i R_i/n) - E(C;n,S) \right]' (S+1)\sum_{i=1}^{k} T_i \right. \\ &\quad \left. - \left[(S+1)\sum_{i=1}^{k}(D_i R_i/n) - E(C;n,S) \right] \left[(S+1)\sum_{i=1}^{k} T_i \right]' \right\}, \end{aligned} \tag{A14}$$

among,

$$\begin{aligned} &\left[(S+1)\sum_{i=1}^{k}(D_i R_i/n) - E(C;n,S) \right]' \\ &= \left\{ (S+1)\sum_{i=1}^{k}(D_i R_i/n) - \left\{ C_o + C_d S \int_0^{\sum_{i=1}^{k} T_i} \delta\left[1 - F_1(t_1)\right] dt_1 + C_p S + C_f^1 (S+1) \int_0^{\sum_{i=1}^{k} T_i} \delta F_1(t_1) dt_1 \right. \right. \\ &\left. + C_f^2 \int_0^{(S+1)\sum_{i=1}^{k} T_i} \lambda_2(t_2) dt_2 \right\} - \left\{ (S+1)\left(\sum_{i=1}^{k} T_i\right) \left[\sum_{j=1}^{k} \left(p_j - D_j \Big/ \left(n \sum_{i=1}^{k} T_i \right) \right) T_j C_h^j \right] \Big/ 2 \right\} - \left. \left\{ (S+1)\sum_{i=1}^{k} C_s^i \right\} \right\}' \\ &= \left\{ -n^{-2}(S+1)\sum_{i=1}^{k} D_i R_i - \left\{ -n^{-1} C_d S \delta\left(\sum_{i=1}^{k} T_i\right)\left[1 - F_1\left(\sum_{i=1}^{k} T_i\right)\right] - n^{-1} C_f^1 (S+1)\delta\left(\sum_{i=1}^{k} T_i\right) F_1\left(\sum_{i=1}^{k} T_i\right) \right. \right. \\ &\left. \left. - n^{-1} C_f^2 (S+1)\left(\sum_{i=1}^{k} T_i\right) \lambda_2\left((S+1)\sum_{i=1}^{k} T_i\right) \right\} - \left\{ -n^{-1}(S+1)\left(\sum_{i=1}^{k} T_i\right) \left[\sum_{j=1}^{k} \left(p_j - D_j \Big/ \left(\sum_{i=1}^{k} D_i/p_i \right) \right) T_j C_h^j \right] \right\} \right\}. \end{aligned} \tag{A15}$$

Then equation (A14) becomes

$$
\begin{aligned}
&E(P;n,S)' \\
&=\Big[\big(S+1\big)\textstyle\sum_{i=1}^{k}T_i\Big]^{-2} \\
&\times\Big\{-n^{-2}\big(S+1\big)\textstyle\sum_{i=1}^{k}D_iR_i-\Big\{-n^{-1}C_dS\delta\Big(\sum_{i=1}^{k}T_i\Big)\Big[1-F_1\Big(\sum_{i=1}^{k}T_i\Big)\Big]-n^{-1}C_f^1\big(S+1\big)\delta\Big(\sum_{i=1}^{k}T_i\Big)F_1\Big(\sum_{i=1}^{k}T_i\Big) \\
&-n^{-1}C_f^2\big(S+1\big)\Big(\textstyle\sum_{i=1}^{k}T_i\Big)\lambda_2\Big(\big(S+1\big)\sum_{i=1}^{k}T_i\Big)\Big\} \\
&-\Big\{-n^{-1}\big(S+1\big)\Big(\textstyle\sum_{i=1}^{k}T_i\Big)\Big[\sum_{j=1}^{k}\Big(p_j-D_j\Big/\Big(\sum_{i=1}^{k}D_i/p_i\Big)\Big)T_jC_h^j\Big]\Big\}\big(S+1\big)\sum_{i=1}^{k}T_i \\
&-\Big\{\big(S+1\big)\textstyle\sum_{i=1}^{k}\big(D_iR_i/n\big)-\Big\{C_o+C_dS\int_0^{\sum_{i=1}^{k}T_i}\delta\big[1-F_1\big(t_1\big)\big]dt_1+C_pS \\
&+C_f^1\big(S+1\big)\textstyle\int_0^{\sum_{i=1}^{k}T_i}\delta F_1\big(t_1\big)dt_1+C_f^2\int_0^{(S+1)\sum_{i=1}^{k}T_i}\lambda_2\big(t_2\big)dt_2\Big\} \\
&-\Big\{\big(S+1\big)\Big(\textstyle\sum_{i=1}^{k}T_i\Big)\Big[\sum_{j=1}^{k}\Big(p_j-D_j\Big/\Big(\sum_{i=1}^{k}D_i/p_i\Big)\Big)T_jC_h^j\Big]\Big/2\Big\}-\Big\{\big(S+1\big)\sum_{i=1}^{k}C_s^i\Big\}\Big\}\Big[-n^{-1}\big(S+1\big)\sum_{i=1}^{k}T_i\Big]\Big\} \\
&=\Big[\big(S+1\big)\textstyle\sum_{i=1}^{k}T_i\Big]^{-2}\Big[-n^{-1}\big(S+1\big)\sum_{i=1}^{k}T_i\Big] \\
&\times\Big\{\Big\{n^{-1}\big(S+1\big)\textstyle\sum_{i=1}^{k}D_iR_i-\Big\{C_dS\delta\Big(\sum_{i=1}^{k}T_i\Big)\Big[1-F_1\Big(\sum_{i=1}^{k}T_i\Big)\Big]+C_f^1\big(S+1\big)\delta\Big(\sum_{i=1}^{k}T_i\Big)F_1\Big(\sum_{i=1}^{k}T_i\Big) \\
&+C_f^2\big(S+1\big)\Big(\textstyle\sum_{i=1}^{k}T_i\Big)\lambda_2\Big(\big(S+1\big)\sum_{i=1}^{k}T_i\Big)\Big\}-\Big\{\big(S+1\big)\Big(\sum_{i=1}^{k}T_i\Big)\Big[\sum_{j=1}^{k}\Big(p_j-D_j\Big/\Big(\sum_{i=1}^{k}D_i/p_i\Big)\Big)T_jC_h^j\Big]\Big\} \\
&-\Big\{\big(S+1\big)\textstyle\sum_{i=1}^{k}\big(D_iR_i/n\big)-\Big\{C_o+C_dS\delta\int_0^{\sum_{i=1}^{k}T_i}\big[1-F_1\big(t_1\big)\big]dt_1+C_pS+C_f^1\big(S+1\big)\delta\int_0^{\sum_{i=1}^{k}T_i}F_1\big(t_1\big)dt_1 \\
&+C_f^2\textstyle\int_0^{(S+1)\sum_{i=1}^{k}T_i}\lambda_2\big(t_2\big)dt_2\Big\}-\Big\{\big(S+1\big)\Big(\sum_{i=1}^{k}T_i\Big)\Big[\sum_{j=1}^{k}\Big(p_j-D_j\Big/\Big(\sum_{i=1}^{k}D_i/p_i\Big)\Big)T_jC_h^j\Big]\Big/2\Big\}-\Big\{\big(S+1\big)\sum_{i=1}^{k}C_s^i\Big\}\Big\}\Big\},
\end{aligned}
\tag{A16}
$$

that is,

$$
\begin{aligned}
&E(P;n,S)' \\
&=\Big[-n\big(S+1\big)\textstyle\sum_{i=1}^{k}T_i\Big]^{-1} \\
&\times\Big\{C_o+C_pS+\Big\{\big(S+1\big)\textstyle\sum_{i=1}^{k}C_s^i\Big\}+C_dS\delta\Big\{\int_0^{\sum_{i=1}^{k}T_i}\big[1-F_1\big(t_1\big)\big]dt_1-\Big(\sum_{i=1}^{k}T_i\Big)\Big[1-F_1\Big(\sum_{i=1}^{k}T_i\Big)\Big]\Big\} \\
&+C_f^1\big(S+1\big)\delta\Big\{\textstyle\int_0^{\sum_{i=1}^{k}T_i}F_1\big(t_1\big)dt_1-\Big(\sum_{i=1}^{k}T_i\Big)F_1\Big(\sum_{i=1}^{k}T_i\Big)\Big\} \\
&+C_f^2\Big\{\textstyle\int_0^{(S+1)\sum_{i=1}^{k}T_i}\lambda_2\big(t_2\big)dt_2-\big(S+1\big)\Big(\sum_{i=1}^{k}T_i\Big)\lambda_2\Big(\big(S+1\big)\sum_{i=1}^{k}T_i\Big)\Big\} \\
&-\Big\{\big(S+1\big)\Big(\textstyle\sum_{i=1}^{k}T_i\Big)\Big[\sum_{j=1}^{k}\Big(p_j-D_j\Big/\Big(\sum_{i=1}^{k}D_i/p_i\Big)\Big)T_jC_h^j\Big]\Big/2\Big\}\Big\}.
\end{aligned}
\tag{A17}
$$

Simplifying equation (A17), we have

$$E(P;n,S)' = \left[n(S+1)\sum_{i=1}^{k} T_i \right]^{-1} \left\{ M(P;n) - \left[C_o + C_p S + (S+1)\sum_{i=1}^{k} C_s^i \right] \right\}, \tag{A18}$$

where

$$M(P;n) = M_1(P;n) + M_2(P;n) + M_3(P;n), \tag{A19}$$

and

$$\begin{cases} M_1(P;n) = \left(C_f^1 (S+1)\delta - C_d S\delta \right) \left[\left(\sum_{i=1}^{k} T_i \right) F_1\left(\sum_{i=1}^{k} T_i \right) - \int_0^{\sum_{i=1}^{k} T_i} F_1(t_1)\, dt_1 \right], \\ M_2(P;n) = C_f^2 \left[(S+1)\left(\sum_{i=1}^{k} T_i \right) \lambda_2\left((S+1)\sum_{i=1}^{k} T_i \right) - \int_0^{(S+1)\sum_{i=1}^{k} T_i} \lambda_2(t_2)\, dt_2 \right], \\ M_3(P;n) = (S+1)\left[\left(\sum_{i=1}^{k} T_i \right) \sum_{j=1}^{k} \left(p_j - D_j \Big/ \left(\sum_{i=1}^{k} D_i / p_i \right) \right) T_j C_h^j / 2 \right], \end{cases} \tag{A20}$$

where $T_i = D_i / (np_i)$.

Since $T_i = D_i / (np_i)$, $\sum_{i=1}^{k} T_i$ strictly decreases with n. In addition, as $F_1(t_1)$ is a strictly non-decreasing function and $\left(\sum_{i=1}^{k} T_i \right) F_1\left(\sum_{i=1}^{k} T_i \right) - \int_0^{\sum_{i=1}^{k} T_i} F_1(t_1)\, dt_1 > 0$, it can be obtained that $\left(\sum_{i=1}^{k} T_i \right) F_1\left(\sum_{i=1}^{k} T_i \right) - \int_0^{\sum_{i=1}^{k} T_i} F_1(t_1)\, dt_1$ generally decreases with n from $\lim_{n\to 0}\left[\left(\sum_{i=1}^{k} T_i \right) F_1\left(\sum_{i=1}^{k} T_i \right) - \int_0^{\sum_{i=1}^{k} T_i} F_1(t_1)\, dt_1 \right]$ and to its lower limit 0. Besides, as the cost for a failure is usually higher than the cost for a detected infect in practice, we assume $C_f^1 > C_d$ and then we have $C_f^1 (S+1)\delta - C_d S\delta > 0$. Therefore, $M_1(P;n)$ generally decreases with n. Similarly, $M_2(P;n)$ also strictly generally decreases with n. Moreover, it is easy to see that $M_3(P;n)$ strictly decreases with n. Thus, $M(P;n)$ is a continuously decreasing function of n.

As $\lim_{n\to 0} M_1(P;n) > 0$, $\lim_{n\to 0} M_2(P;n) > 0$ and $\lim_{n\to 0} M_3(P;n) \to \infty$, it is easy to know that $\lim_{n\to 0} M(P;n) = \infty$. On the contrary, we have $\lim_{n\to\infty} M(P;n) = 0$, since $\lim_{n\to\infty} M_1(P;n) = 0$, $\lim_{n\to\infty} M_2(P;n) = 0$ and $\lim_{n\to\infty} M_3(P;n) = 0$. Therefore, as $C_o + C_p S + (S+1)\sum_{i=1}^{k} C_s^i > 0$, there exists a finite and unique n^* which satisfies $M(P;n^*) - \left[C_o + C_p S + (S+1)\sum_{i=1}^{k} C_s^i \right] = 0$ and further makes $E(P;n^*,S)' = 0$. Besides, we also have $\lim_{n\to 0} E(P;n,S)' > 0$ and $\lim_{n\to\infty} E(P;n,S)' < 0$, so $E(P;n^*,S)$ is the maximum of $E(P;n,S)$ and $E(P;n,S)$ is a strictly concave function of n for any $S > 0$. n^* can be obtained by setting equation (A18) to be 0.

This establishes **Proposition 3**.