

12-31-1994

Timing of Software Replacement

Taizan Chan
National University of Singapore

S. Chung
University of Calgary

Follow this and additional works at: <http://aisel.aisnet.org/icis1994>

Recommended Citation

Chan, Taizan and Chung, S., "Timing of Software Replacement" (1994). *ICIS 1994 Proceedings*. 70.
<http://aisel.aisnet.org/icis1994/70>

This material is brought to you by the International Conference on Information Systems (ICIS) at AIS Electronic Library (AISEL). It has been accepted for inclusion in ICIS 1994 Proceedings by an authorized administrator of AIS Electronic Library (AISEL). For more information, please contact elibrary@aisnet.org.

TIMING OF SOFTWARE REPLACEMENT

Taizan Chan
Siu Leung Chung
Teck Hua Ho

Department of Information Systems and Computer Science
National University of Singapore

ABSTRACT

Analogous to the replacement of an old machine, such as a car, replacing an aged software may contain its escalating cost of maintenance. Prior research has assumed that increasing maintenance cost is due to the deterioration of the system maintainability. However, cost of maintenance depends also on the number of incoming maintenance requests. While software maintainability is determined by its complexity and development environment, number of maintenance requests is affected by the business environment. This distinction is significant in analyzing economic tradeoffs in software maintenance because replacement improves system maintainability but will not affect the number of maintenance requests. Unlike replacement of hardware, rewriting software takes an extended length of time. Thus, the old software must still be maintained before the new software is ready. We develop an economic model that considers the number of maintenance requests and the rewriting period explicitly. The model is an extension of Gode, Barua, and Mukhopadhyay (1990), which assumes a constant number of maintenance requests and instantaneous replacement. Our model allows us to draw some additional policy implications about software maintenance and replacement. For instance, we show that in certain situations delaying a system replacement can be more cost effective when the user environment changes more rapidly, contrary to our intuition. Moreover, it is shown that rewriting should begin earlier when the instantaneous replacement assumption is relaxed.

1. INTRODUCTION

The cost of maintaining application software has been rapidly increasing. It is currently estimated to comprise 50% to 80% of the corporate software budgets in the United States (Banker, Datar, and Kemerer 1991; Gode, Barua, and Mukhopadhyay 1990). This cost is enormous when the expenditure on software in the United States *alone* is estimated at \$100 billion. Despite this, economics of software maintenance has been a relatively neglected area of research (Bakos and Kemerer 1992). As the trend of increasing maintenance cost is likely to continue in the foreseeable future, it is important that more research effort be directed at the study of the economics of software maintenance and the analysis of means to control this huge expense.

According to Swanson (1976), software maintenance can be classified into 1) adaptive, 2) perfective, and 3) corrective maintenance. In our study, we focus on adaptive and perfective maintenance, which was estimated to occupy as high as 75% of the maintenance work (Lientz, Swanson,

and Thompkins 1978), and call them enhancements. Due to rapidly changing business environments and user needs, a system has to be continuously modified and enhanced. With frequent modifications and enhancements, the number of functions, control flows, and inter-module interactions in the system increase over time, leading to higher complexity. Moreover, these enhancements are normally not well integrated in the overall design and not well documented, leading to deteriorating system structure and poor maintainability. As a result, the effort required for maintenance increases sharply due to the degradation of the system (Banker, et al. 1993; Gibson and Senn 1989; Jones 1989). Thus, there may exist a time when it is economically justified to rewrite the system, which may result in the improvement of system maintainability and hence the reduction in the cost of maintenance over a planning horizon (Gode, Barua, Mukhopadhyay 1990). Swanson and Beath (1989) have found that replacement of old systems is in fact a significant activity, thereby emphasizing the importance of studying the economic ramifications of software maintenance and replacement policies.

Gode, Barua, Mukhopadhyay have analyzed the economic tradeoffs involved in software replacement. They suggested that replacement of an aged software can be useful and provided a model for determining the optimal rewriting point(s). They assumed that costs of maintenance are convex functions of the cumulative amount of maintenance done, measured in units of maintenance such as Function Points (see Albrecht and Gaffney [1983] for an introduction on function point measurements). They showed that the optimal rewriting time can be influenced by the initial size of the system and the structuredness of the underlying technologies of the old and new systems. This work is the first model to formally analyze the economic tradeoffs in software replacement. It provides us with managerial guidelines for the formulation of rewriting policies in terms of system maintainability. The model, however, makes two simplifying assumptions. First, it assumes that the number of incoming maintenance requests is uniform over time. Second, the replacement of software is assumed to occur instantaneously. We relax both assumptions in this paper.

The maintenance effort of an application at a particular time is the product of the number of maintenance requests and the effort needed per maintenance request. Figure 1 shows that the number of maintenance request is a business-related factor while the effort per maintenance request is a system-related factor. An important implication of this observation is that rewriting the system affects only the maintenance effort per request but does not alter the pattern of the maintenance requests. In addition, we can now study the impact of the pattern of maintenance requests on software replacement policies.

Rewriting a complex application can take an extended length of time. During rewriting, the old system must still be maintained. Depending on the software development methodology (Davis, Bersoff, and Comer 1988), the design specification of the new system may also lag the current user needs. These additional costs must be incorporated in a comprehensive model that considers the rewriting period explicitly.

The remainder of the paper is organized as follows: section 2 describes the model framework; section 3 considers instantaneous replacement; section 4 examines the impact of a finite rewriting period; section 5 discusses the results and suggests future research directions.

2. MODEL FRAMEWORK

We consider the cumulative maintenance effort of an independent application over a planning horizon T . The planning horizon starts at the time when the application is operational and ends at the time when it is obsolete.¹ It is assumed that the application manager receives and fulfills

an incoming stream of maintenance requests from users. The system maintainability deteriorates as more and more modifications and enhancements are made on the system. At T_R , the application manager begins the development of a new system whose functionality is equivalent to the old system at T_R . This development is scheduled to end at T_N , the time when the old system is withdrawn and the new system is operational. Figure 2 shows the problem scenario we wish to model here.

In the following subsections, we describe the forces that determine the number of maintenance requests and the effort per maintenance request. We then formulate the maintenance planning problem as a constrained non-linear optimization problem.

2.1 Business-Related Factor: Number of Maintenance Requests

The number of incoming maintenance requests is driven by how rapidly a business environment changes (Martin and McClure 1983). As business scenarios change, the application must be modified or enhanced. These changes give rise to requests for maintenance over time. A more fluid environment will generate a higher number of requests and hence require a higher cumulative effort for maintenance. Note that the number of maintenance requests is independent of the system maintainability. Let $m(t)$ be the number of maintenance requests at time t . One way to represent the growth of a variable commonly used in economic analysis is to represent the number of maintenance requests at time t as

$$m(t) = m(0)e^{\delta t} \quad (2.1)$$

where $m(0)$ is the number of maintenance requests at the time the system first becomes operational and $\delta > 0$ is the logarithmic rate of growth (see Figure 3).

Note that these parameters can be estimated reliably by regression analysis with data relating to the distribution of maintenance requests over time. A high $m(0)$ implies a large user base. A high δ reflects a volatile business environment in which system requirements increase rapidly with time. A low δ depicts a stable environment. It is important to note that replacement *does not* alter the number of maintenance requests over time. The replacement, however, will improve the system maintainability which in turn lowers the maintenance effort per request (described below). In this paper, we assume that all maintenance requests are fulfilled (that is, we have no backlog of requests.) We discuss how capturing the number of maintenance requests can open up new research opportunities in modeling backlogs in a loaded information system department in section 5.

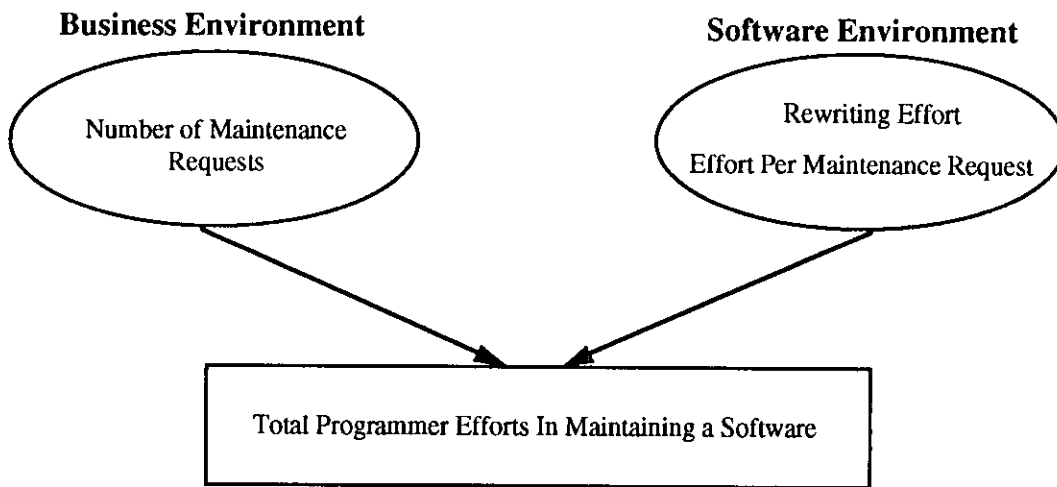


Figure 1. The Business- and System-Driven Factors Affecting Total Maintenance Effort

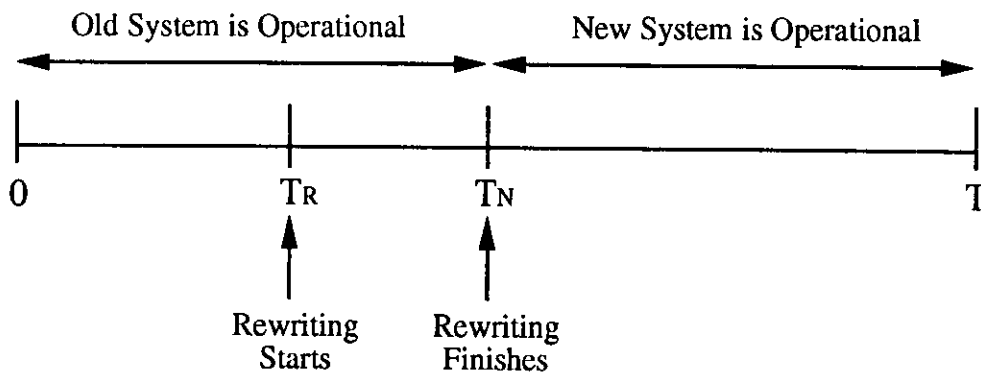


Figure 2. The Problem Scenario

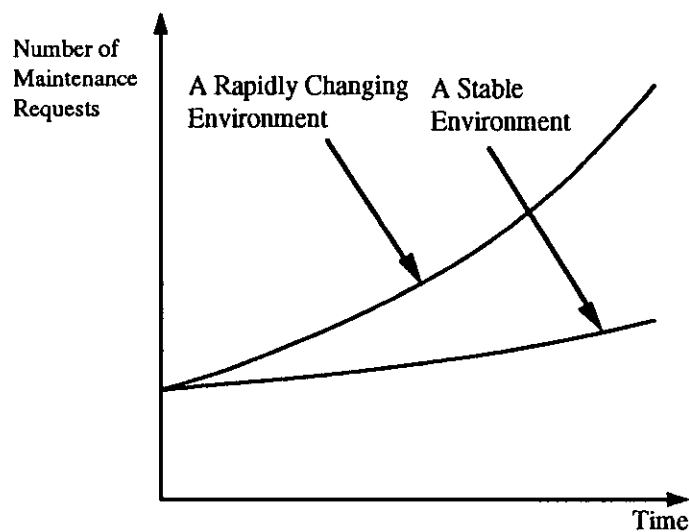


Figure 3. Patterns of Maintenance Requests in Two Business Environments

2.2 System-Related Factor: Effort Per Maintenance Request

The effort per maintenance request is a function of the system maintainability. Similar to Gode, Barua, and Mukhopadhyay, we assume that system maintainability is directly related to the cumulative number of maintenance jobs that have been performed on the system. The cumulative number of maintenance jobs at time t , $M(t)$, is the integral sum of the number of maintenance requests during the interval $[0, t]$. That is,

$$M(t) = \int_0^t m(t)dt.$$

If $m(t)$ is exponential (see above), $M(t)$ can be expressed as

$$M(t) = \frac{m(0)}{\delta}(e^{\delta t} - 1). \quad (2.2)$$

Let $h_o(\cdot)$ and $h_n(\cdot)$ be the effort required for a maintenance job on the old and new systems respectively. If the system is rewritten with the same technology, then $h_n(\cdot)$ equals $h_o(\cdot)$ if their system maintainability is identical. If the system is rewritten with a superior technology that improves system maintainability, then $h_n(\cdot)$ is smaller than $h_o(\cdot)$.

2.3 The Maintenance Planning Problem

During the planning horizon, three types of effort are expended. First, effort is needed to maintain the old system. The total effort required for the maintenance of the old system is

$$\int_0^{T_N} m(t)h_o(M(t))dt. \quad (2.3)$$

That is, the total effort required for maintaining the old system is the sum of all the effort required for fulfilling all the maintenance requests from the time the old system starts operating to the time at which the old system is withdrawn.

Second, effort is needed to maintain the new system. Similarly, the total effort required for maintenance of the new system is

$$\int_{T_R}^T m(t)h_n(M(t) - M(T_R))dt \quad \text{or} \\ \int_0^{T-T_R} m(t+T_R)h_n(M(t+T_R) - M(T_R))dt. \quad (2.4)$$

That is, the total effort required for maintaining the new system is the integral sum of all the effort required for fulfilling all the maintenance requests from the time the design specification is "frozen" up to the end of the planning horizon, T . Two points to note: First, the maintenance period for the new system is $T - T_R$ instead of $T - T_N$. This is to account for any enhancement requests that arrive during the rewriting period. Second, the number of maintenance requests at T_R is equal to $m(T_R)$ and not $m(0)$ because it is not affected by the rewriting and system maintainability.

The third effort to be expensed is the rewriting effort for the new system. A general formulation of this effort is

$$L(T_N - T_R), \quad (2.5)$$

where L is the size of the development team that is employed to rewrite the system. The rewriting effort is related to the complexity of the old system at the time of rewriting. The complexity of a system is usually expressed in terms of the total number of Function Points the system has (Albrecht and Gaffney 1983). The complexity of the old system at time T_R is the sum of the original complexity of the system, θ_0 (when it was first installed) and the added complexity due to maintenance. The added complexity is dependent on the complexity of each maintenance request over the planning horizon. If average complexity of a maintenance request is θ_m , then the complexity of the application at time t , $F(M(t))$ is given by

$$F(M(t)) = \theta_0 + \theta_m M(t). \quad (2.6)$$

We conceptualize system development as a Function Point production process (Banker, Datar, and Kemerer 1991). The number of Function Points produced per period is a log-linear form of the development team size, i.e., $\omega_0 L^{\omega_1}$. ω_0 measures the structuredness of the development technology used and ω_1 measures the productivity of the rewriting team. A high ω_0 implies a highly structured development technology. A high ω_1 implies a rewriting team whose members are very experienced and are familiar with the system under maintenance. In most situations, $\omega_1 < 1$, which implies a diminishing return to labor input. Since the rewriting period is $(T_N - T_R)$, we have

$$F(M(T_R)) = \omega_0 L^{\omega_1} (T_N - T_R). \quad (2.7)$$

The total cumulative effort of maintaining the application over the planning horizon T , $E(T_R, T_N)$, is the sum of the three components, (2.3), (2.4), and (2.5):

$$E(T_R, T_N) = \int_0^{T_N} m(t)h_o(M(t))dt + \int_0^{T-T_R} m(t+T_R)h_n(M(t+T_R)-M(T_R))dt + L(T_N - T_R). \quad (2.8)$$

Thus, the optimization problem is

$$[G] \quad \min_{T_R, T_N} E(T_R, T_N) \\ \text{subject to } F(M(T_R)) = \omega_0 L^{\omega_1} (T_N - T_R), \\ \text{and } T_R \leq T_N. \quad (2.9)$$

In Section 3, we solve problem [G] with $\omega_1 = 1$. $\omega_1 = 1$ implies that there is no penalty associated with compressing the rewriting schedule. The optimal replacement policies suggest that, in this case, instantaneous replacement is optimal, an assumption of the Gode, Barua, and Mukhopadhyay model. This allows us to compare our results with theirs. In section 4, we solve the problem [G] with the size of the rewriting team kept fixed. This enables us to investigate the impact of a finite rewriting period.

3. INSTANTANEOUS REPLACEMENT

If $\omega_1 = 1$, it can be shown that rewriting effort is $(1/\omega_0)F(M(T_R))$ and the total cumulative effort over the planning horizon is given by

$$E(T_R, T_N) = \int_0^{T_N} m(t)h_o(M(t))dt + \int_0^{T-T_R} m(t+T_R)h_n(M(t+T_R)-M(T_R))dt + \frac{F(M(T_R))}{\omega_0}. \quad (3.1)$$

The optimization problem is

$$[G1] \quad \min_{T_R, T_N} E(T_R, T_N) \\ \text{subject to } T_R \leq T_N. \quad (3.2)$$

We consider two maintenance scenarios. The first scenario represents an environment where there is little fixed effort needed for carrying out the maintenance job. However, the variable maintenance effort needed grows rapidly with the cumulative number of maintenance requests so that a log-linear function is an appropriate assumption. This is the common assumption found in the previous literature (Boehm 1981; Gode, Barua, and Mukhopadhyay 1990). The second scenario depicts an environment where a

significant fixed effort is needed to fulfill a maintenance request. This fixed effort may consist of a rigorous decision-making process, elaborate documentation, and standard trial runs using real data. This extra "preventive" effort (Martin and McClure 1983) results in a slower rate of increase in maintenance effort with respect to the cumulative number of maintenance requests. We approximate this environment with a linear function.

3.1 Case 1: The Log-Linear Case

The log-linear assumption may be suitable for modeling some traditional system environments. First, it is appropriate to describe a system that had been developed with old or unstructured technology, such as COBOL. This usually leads to a rapid increase in the effort for maintenance since the system maintainability deteriorates significantly as modifications and enhancements are performed on the system. The next suitable environment is one in which there is little or no documentation done. An outdated documentation implies a greater effort for new maintenance requests since the programmers must examine the codes to resolve any inconsistencies. This extra effort rises rapidly as the system becomes more complicated and the documentation more outdated.

In the log-linear case, h_o and h_n take the following forms:

$$h_o(M(t)) = \alpha_M^o M(t)^{\beta_M^o}, \quad (3.3)$$

$$h_n(M(t)) = \alpha_M^n M(t)^{\beta_M^n}. \quad (3.4)$$

α_M^o and α_M^n measure the structuredness of the underlying technology for the old and new systems respectively. A high α_M means that the structuredness of the maintenance environment is poor. β_M^o and β_M^n reflect the productivities of the maintenance team for the old and new systems respectively. A higher β_M means a less productive maintenance team.²

Solving problem [G1], we obtain the following proposition:

Proposition 1. If $\delta < \ln(\beta_M^n + 1)/T$, solution to [G1] exists and is unique. At optimal $T_R^* = T_N^*$ and T_R^* is characterized by

$$\alpha_M^n \left[e^{\delta T_R^*} M(T - T_R^*) \right]^{\beta_M^n} = \alpha_M^o M(T_R^*)^{\beta_M^o} + \frac{\theta_m}{\omega_0} \quad (3.5)$$

Proof. See the appendix.

Note that the optimal policies prescribe an instantaneous replacement if $\omega_1 = 1$. Based on the above proposition, we derive several managerial insights relating the optimal replacement policy to the business and system development environment, as depicted in Figures 4a through 4f. The first three are new insights: the first relates to the rate of growth of number of maintenance requests, the second to the initial system size, and the third to the average complexity of maintenance request. The last three results are consistent with those of Gode, Barua, and Mukhopadhyay. See the appendix for proofs of these managerial implications.

1. T_R^* increases with the rate of growth of number of maintenance requests (δ). That is, the more volatile the business environment is, the later should the old system be replaced. This observation implies that replacement should be done later in a business environment in which user needs change rapidly with time. This result seems counter-intuitive: in order to reduce the escalating maintenance effort of the old system, one may expect to replace the old system earlier to improve its rapidly deteriorating system maintainability. The result can be understood better if we compare a stationary and a changing environment. Compared to the stationary environment, the changing environment has more maintenance jobs to be done in the later part of the planning horizon. That is, the system maintainability will deteriorate more dramatically and the effort per maintenance request will be higher during that period. This higher effort per request "pulls" the replacement timing to a later date so that the effort per request of the old and new systems are more "balanced."³
2. T_R^* is independent of the initial system size (θ_0). This result is in contrary to that of Gode, Barua, and Mukhopadhyay. They found that the operational life of a system decreases with an increase in the initial system size. Our model reveals that the initial system size represents a "fixed" effort required for rewriting if a constant return to labor input is assumed (i.e., $\omega_1 = 1$). We shall show in section 4 that, if there is diminishing return to labor input ($\omega_1 < 1$), T_R^* decreases with the initial system size.
3. T_R^* decreases with the average complexity of maintenance request (θ_m). That is, the greater the average complexity, the earlier should the system be replaced. This is because an increase in θ_m implies an increase in $F(M(T_R))$ and thus an increase in the marginal rewriting effort with respect to T_R . Thus, the system should be replaced earlier to reduce this effort.
4. T_R^* increases with the productivity of the development technology (ω_0). That is, the more superior the devel-

opment technology, the later can the replacement be delayed. The result implies that the system manager must start rewriting earlier if an existing and inferior technology is used in preference to a costly and superior technology. The cost of a superior technology can then be assessed against its potential savings in the cumulative effort of maintenance over a planning horizon. Thus, a lack of advanced tools implies an earlier rewriting time.

5. T_R^* decreases when the structuredness of the new technology is more superior or the structuredness of the old technology is more inferior (i.e., a decrease in α_M^o or an increase in α_M^o .) A lower α_M^o or a higher α_M^o implies a better new technology that reduces the effort required for each maintenance request for the new system. This means that a system manager should consider replacing the system earlier if it has been decided that a superior technology will be adopted for the new system.
6. T_R^* decreases when the maintenance team for new system is more productive or the maintenance team for the old system is less productive (i.e., a decrease in β_M^n or an increase in β_M^o .) A lower β_M^n or a higher β_M^o implies a more productive maintenance team for the new system. This is the case if the staff assigned to maintaining the new system is familiar with the technology used for rewriting the new system. In this case, the maintenance team is likely to be composed of the staff who are also involved in developing the new system. This would imply less learning overhead and consequently better productivity. Thus, replacement can start earlier if the system manager decides to organize the IS staff according to application systems (Swanson and Beath 1989) in contrast to, say, separating the staff into development and maintenance teams.

3.2 Case 2: The Linear Case

The linear case assumes that the programmer effort is constant to scale. This may describe a situation in which the IS staff is organized according to application skills. That is, an IS staff responsible for the development of a system will also be responsible for its maintenance. Thus, the maintenance staff will be very familiar with the system such that little overhead (such as understanding codes) is incurred for each maintenance request. The linear functions are also suitable for modeling a stringent, well-controlled system development environment. In such an environment, each maintenance request is subjected to a rigorous approval process, documentation, and testing. We assume the following functional forms for h_o and h_n .

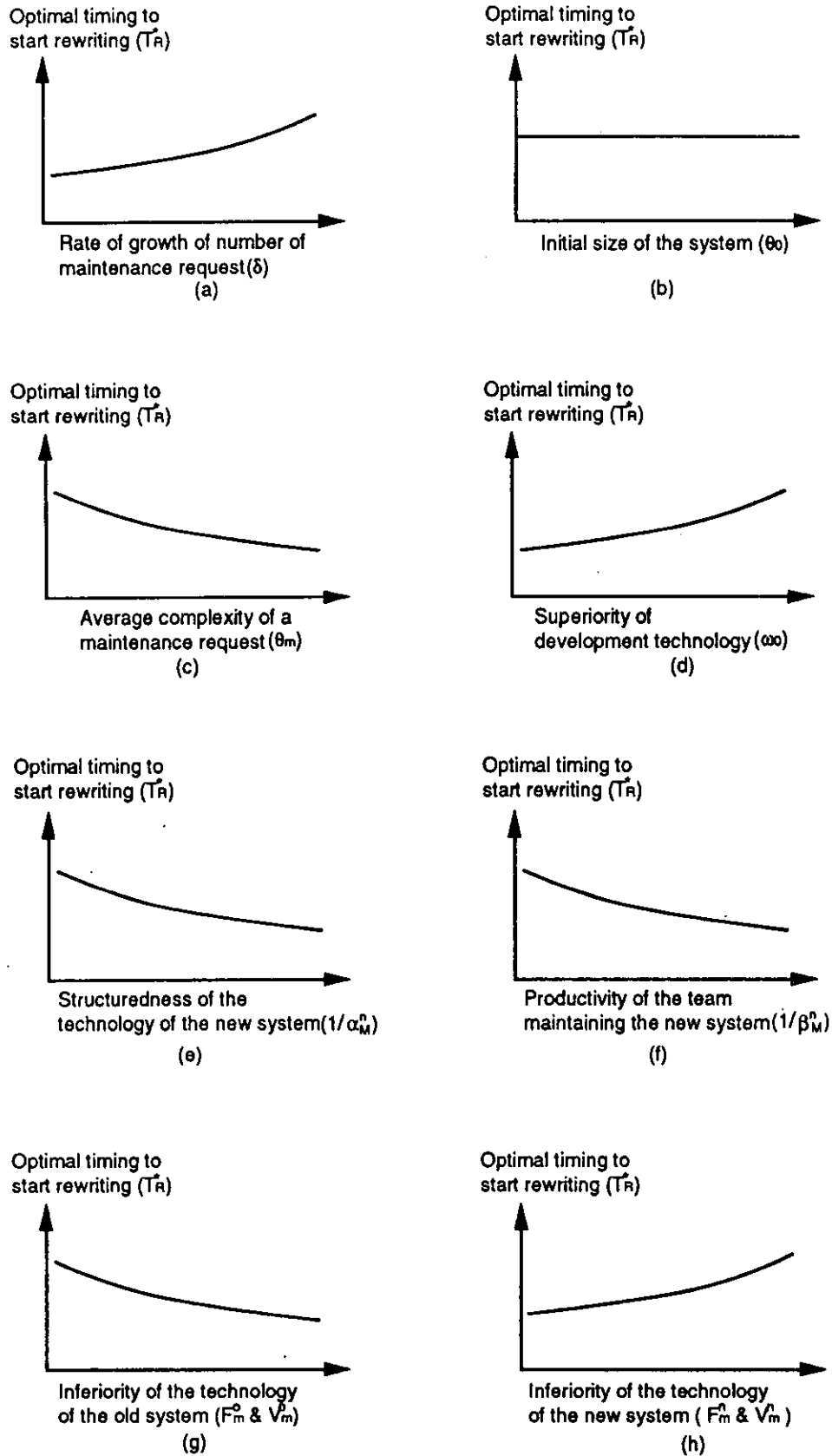


Figure 4. Implications for Software Replacement when $\omega_1 = 1$

$$h_o(M(t)) = F_M^o + V_M^o M(t), \quad (3.6)$$

$$h_n(M(t)) = F_M^n + V_M^n M(t). \quad (3.7)$$

F_M^o and F_M^n measure the fixed effort required for each maintenance request for the old and new systems respectively. Note that this component of the maintenance effort is independent of the system maintainability. A high F_M^o or F_M^n means a tight control over the maintenance process. For instance, each maintenance request may be subjected to a standard approval process, paper work, and testing procedures. F_M^o and F_M^n can be different because a more superior technology can reduce the fixed effort through some semi-automatic assistance. V_M^o and V_M^n are the constants of proportionality for the variable part of the maintenance effort for the old and new systems respectively. They are similar to the α 's in the log-linear case in that they measure the structuredness of the maintenance technology.

Solving the minimization problem [G1] gives us the following proposition:

Proposition 2. If $F_M^n < \theta_m/\omega_0$ and $\delta < (\ln 2)/T$, solution to [G1] exists and is unique. At optimal, $T_R^* = T_N^*$ and T_R^* is characterized by

$$0 \frac{m(0)V_M^n}{\delta} (e^{\delta T} - e^{\delta T_R^*}) + F_M^n = \quad (3.8)$$

$$\frac{m(0)V_M^o}{\delta} (e^{\delta T_R^*} - 1) + F_M^o + \frac{\theta_m}{\omega_0}.$$

Proof. See the appendix.

The above proposition also suggests that T_R^* is independent of the initial system size (θ_0). It decreases with the average complexity of maintenance requests (θ_m) and increases with the rate of growth of number of maintenance requests (δ) and the structuredness of the development technology (ω_0). In addition, it yields two managerial insights depicted in Figures 4g and 4h (see the appendix for proof). These insights are consistent with those of Gode, Barua, and Mukhopadhyay. Figures 4g and 4h show that T_R^* decreases with F_M^o and V_M^o , and increases with F_M^n and V_M^n . That is, the more inferior the old technology is compared to the new, the earlier we want to replace the system to take

advantage of the savings in effort that the new technology offers.

4. IMPACT OF FINITE REWRITING PERIOD

We consider another case of the general problem [G], where the firm is constrained by limited programming resources. This is found to be common in most information system departments (Swanson and Beath 1989). In particular, we consider the case where the size of development team is fixed at L_R . If $L = L_R$, it can be shown that the total cumulative effort over the planning horizon is given by

$$E(T_R, T_N) = \int_0^{T_N} m(t)h_o(M(t))dt + \int_0^{T-T_R} m(t+T_R)h_n(M(t+T_R)-M(T_R))dt + \frac{F(M(T_R))L_R^{1-\omega_1}}{\omega_0}. \quad (4.1)$$

The optimization problem is

$$[G2] \quad \min_{T_R, T_N} E(T_R, T_N) \quad (4.2)$$

subject to $T_N = T_R + \frac{F(M(T_R))}{\omega_0 L_R^{\omega_1}}$

We derive the optimal software replacement policies with respect to the time at which rewriting starts, T_R , and the time at which rewriting finishes, T_N . Similarly we will consider both the log-linear and the linear case.

Case 3. Log-Linear Case

Solving [G2], we have the following proposition concerning T_R and T_N :

Proposition 3. If $\delta < \ln(\beta_M^n + 1)/T$, solution to [G2] exists and is unique. At optimal, T_R^* and T_N^* are uniquely characterized by

$$\alpha_M^n \left[\frac{m(0)}{\delta} (e^{\delta T} - e^{\delta T_R^*}) \right]^{\beta_M^n} = \alpha_M^o M(T_N^*)^{\beta_M^o} e^{\delta(T_N^* - T_R^*)} \left[1 + \frac{\theta_m}{\omega_0 L_R^{\omega_1}} m(T_R^*) \right] + \frac{\theta_m}{\omega_0} L_R^{1-\omega_1}. \quad (4.3)$$

Proof. See the appendix.

Case 4. Linear Case.

Solving [G2], we have the following proposition concerning T_R and T_N :

Proposition 4. If $F_M^n < (L_R^{1-\omega_1}\theta_m)/\omega_0$ and $\delta < (\ln 2)/T$, solution to [G2] exists and is unique. At optimal, T_R^* and T_N^* are uniquely determined by

$$F_M^n + \frac{m(0)V_M^n}{\delta}(e^{\delta T_R} - e^{\delta T_N}) = \left[F_M^n + \frac{m(0)V_M^n}{\delta}(e^{\delta T_R} - 1) \right] e^{\delta(T_N - T_R)} + \frac{\theta_m}{\omega_0 F_R^n} m(T_R^*) + \frac{L_R^{1-\omega_1}\theta_m}{\omega_0}(e^{\delta T_R} - 1). \tag{4.4}$$

Proof. See the appendix.

Similar insights can be obtained with respect to θ_m , ω_0 , α_M^o , α_M^n , β_M^o , β_M^n , F_M^o , F_M^n , V_M^o , and V_M^n . In addition, Propositions 3 and 4 suggest the following (see Figure 5):

1. T_R^* increases with productivity of the rewriting team (ω_1). That is, with a more productive rewriting team, the firm can afford to postpone the rewriting to a later date. Note that our T_R^* is smaller than that predicted by the model given by Gode, Barua, and Mukhopadhyay since they assume $\omega_1 = 1$.
2. T_R^* decreases with the initial system size (θ_0). This result is consistent with that of Gode, Barua, and Mukhopadhyay. However, our analysis suggests that this result is only valid when there is a finite rewriting period (see implication 2 from Proposition 1).
3. The rewriting period ($T_N^* - T_R^*$) decreases when the new technology is more superior (a smaller α_M^n or F_M^n or V_M^n). During rewriting, maintenance efforts must still be spent to fulfill maintenance requests for the old system in addition to keeping the new system current. This result implies that a better technology will not only reduce the maintenance efforts for the new system over the planning horizon but also decrease the duplication of maintenance during rewriting.
4. The rewriting period ($T_N^* - T_R^*$) decreases when the maintenance team for the new system is more productive (a smaller β_M^n). This shortening of the rewriting period will again reduce the maintenance effort for the new system over the planning horizon and the duplication of maintenance.

5. DISCUSSION

Gode, Barua, and Mukhopadhyay provide the first attempt to formally analyze the economic tradeoffs in software

replacement. They assume that the main determinant of the rising maintenance cost is the system maintainability and that system replacement is instantaneous. We relax both assumptions. We have made a contribution by providing a much more realistic model. Our model incorporates the "user-side" of the software maintenance problem. We show that it can be optimal to delay the replacement of the old system if the business environment changes rapidly. Our model also considers the impact of a finite rewriting period. We show that an instantaneous replacement is only optimal when we have a constant return to scale in development and an ample development capacity. When these conditions do not hold, some of the insights derived must be modified accordingly. For instance, the initial system size which does not influence the optimal time to replace in the former case is found important in the latter case. Our model yields the following managerial implications:

1. *Delay replacement if the rate of growth of the number of requests is high.* This insight suggests that IS organizations must consider not only present but also future patterns of maintenance requests when deciding on a timing for software replacement. Assuming a constant pattern of requests might lead to a suboptimal replacement decision. For instance, if the rate of growth is positive but is assumed to be zero, this will lead to an earlier replacement. Consequently, the cumulative maintenance efforts will be higher because the new system will undergo many changes and become hard to maintain at the end of the planning horizon. A long-term perspective that examines users' future needs is needed for minimizing such bias.
2. *Assign the most experienced staff to rewriting.* We have shown that, if there is a constant return to scale, the rewriting period diminishes. To reduce the duplication of maintenance during rewriting, an IS organization should assign a productive team to rewriting.⁴ This can be achieved by composing a team of staff who are experienced in using the tools and technology chosen for the development and familiar with the system.
3. *Train the maintenance team with the new technology.* We have shown that the rewriting period decreases if the new technology is more superior and increases if the maintenance team for the new system is less productive. Therefore, if a superior new technology is not accompanied by a productive maintenance team, the rewriting period may not decrease. This is because the potential decrease in rewriting period due to the superiority of the technology will be offset by the increase in rewriting period due to the lower productivity of the maintenance team. To reduce length of rewriting, it is important that the maintenance team be familiar with the new technology and be able to fully exploit it.

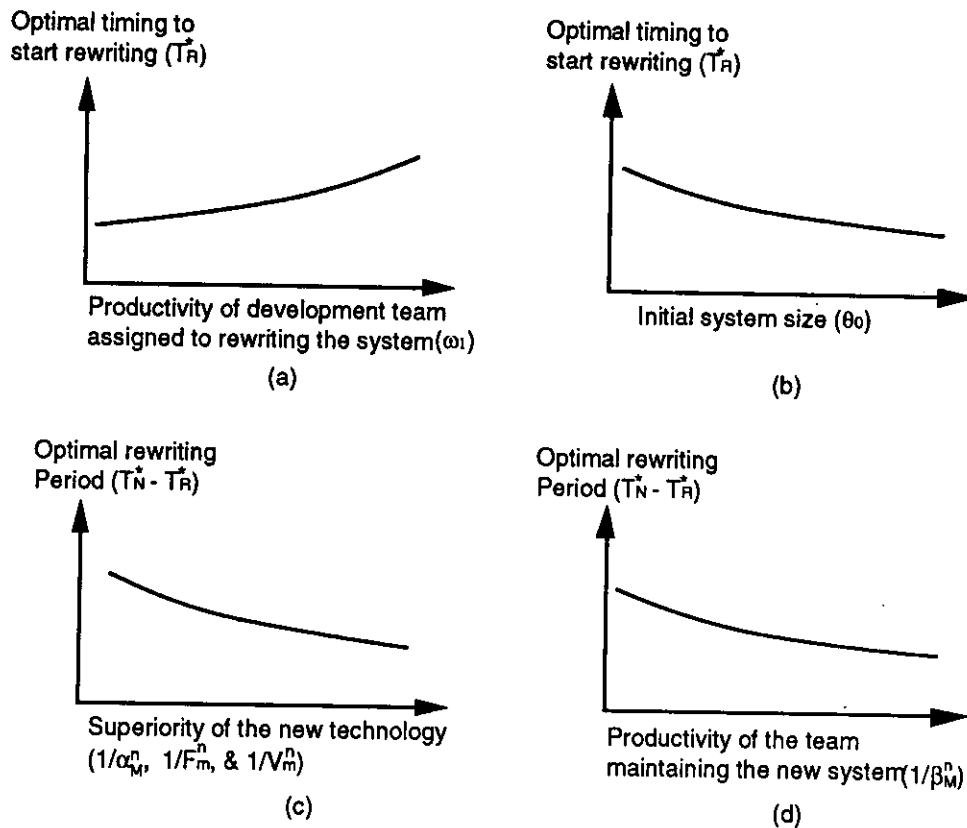


Figure 5. Implications for Software Replacement When There is a Finite Rewriting Period

Our model also opens up several new research possibilities.

1. *Study the effect of maintenance backlogs.* Our analysis has assumed that all maintenance requests are fulfilled. In many situations where the maintenance team size is relatively small compared to the size of the user base, not all of the maintenance requests can be satisfied. Some maintenance requests must be shelved for more urgent ones. By breaking the maintenance request function, $m(t)$, into fulfilled and unfulfilled requests and assigning a penalty cost to each unfulfilled request, we can study how the software replacement policy will be affected by the amount of backlogs.
2. *Adopt a software development life cycle (SDLC) perspective.* We have modeled the maintenance phase of the SDLC. However, it is well recognized that the amount of effort required for maintaining software is related to how well the system has been developed and tested in the earlier phases of the SDLC. Therefore, a

more general model for the economics of software maintenance should adopt a broader perspective of SDLC and determine the interactions between development and maintenance.

3. *Evaluate different policies for assignment of maintenance staff.* Swanson and Beath have reported three different ways in which an IS department organize its staff: by skills (staff is divided into system analysts and programmers), by applications (staff is divided according to applications) and by life-cycle (staff is divided into development and maintenance teams). We hypothesize that different staff organizations imply different levels of maintenance productivity. These differences are captured in our model via different β_M 's. Consequently, the overall maintenance effort and replacement policies could be different. By collecting empirical data relating maintenance effort to staff organizations, we can determine the values of β_M 's and the merits of each organization.

4. *Study interactions between applications in a portfolio.* We have assumed a single application in our model. In many real-life situations, however, a system manager has to wrestle with a portfolio of applications (Swanson and Beath 1989). In these situations, the manager must decide how to allocate a common pool of programmers to different applications. The scarcity of programmer resources gives rise to backlogs and frequent switching of programmers from one system to another. These represent interesting extensions to our model.

6. REFERENCES

Albrecht, A. J., and Gaffney, J. E. "Software Function, Source Lines of Code, and Development Effort Prediction: A Software Science Validation." *IEEE Transactions on Software Engineering*, Volume 9, Number 6, 1983, pp. 639-648.

Bakos, J. Y., and Kemerer, C. F. "Recent Applications of Economic Theory in Information Technology Research." *Decision Support Systems* Volume 8, 1992, pp. 365-386.

Boehm, B. *Software Engineering Economics*. Englewood Cliffs, New Jersey: Prentice-Hall, 1981.

Banker, R. D.; Datar, S. M.; and Kemerer, C. F. "A Model to Evaluate Variables Impacting the Productivity of Software Maintenance Projects." *Management Science*, Volume 37, Number 1, 1991, pp. 1-18.

Banker, R. D.; Datar, S. M.; Kemerer, C. F.; and Zweig, D. "Software Complexity and Maintenance Costs." *Communications of ACM*, Volume 36, Number 11, 1993, pp. 81-94.

Chan, T.; Chung, S.; and Ho, T. H. "Timings of Software Rewriting and Replacement." Submitted to *IEEE Transactions on Software Engineering*, 1994.

Davis, A.; Bersoff, E.; and Comer, E. "A Strategy for Comparing Alternative Software Development Life Cycle Models." *IEEE Transactions on Software Engineering*, Volume 14, Number 10, 1988, pp. 1453-1461.

Gibson, V. R., and Senn, J. A. "System Structure and Software Maintenance Performance." *Communications of the ACM*, Volume 32, Number 3, 1989, pp. 347-358.

Gode, D. K.; Barua, A.; and Mukhopadhyay, T. "On the Economics of the Software Replacement Problem." In J. I. DeGross, M. Alavi, and H. Oppelland (Editors), *Proceedings of the Eleventh International Conference on Information Systems*, Copenhagen, Denmark, 1990, pp. 159-170.

Jones, C. "Software Enhancement Modelling." *Software Maintenance: Research and Practice*, Volume 1, 1989, pp. 91-100.

Lientz, B. P.; Swanson, E. B.; and Tompkins, G. E. "Characteristics of Application Software Maintenance." *Communications of the ACM*, Volume 21, Number 6, 1978, pp. 466-471.

Martin, J., and McClure, C. *Software Maintenance: The Problem and Its Solution*. Englewood Cliffs, New Jersey: Prentice-Hall, 1983.

Swanson, E. B. "The Dimensions of Maintenance." In *Proceedings of the Second International Conference on Software Engineering*, 1976, pp. 492-497.

Swanson, E. B., and Beath, C. M. *Maintaining Information Systems in Organizations*. New York: John Wiley and Sons, 1989.

7. ENDNOTES

1. We make a distinction between application and system. The old system is obsolete when it is withdrawn and the application is obsolete at the end of the planning horizon.
2. It is worthwhile noting that having a low α_M does not necessarily imply a low β_M . The maintenance effort for the new system will be lower only if a better technology is coupled with a team of staff who can work with the technology efficiently (cf., Gode, Barua, and Mukhopadhyay 1990).
3. Note that we ignore the time value of money. Incorporating a discount factor will tilt the balance to the left of the planning horizon (i.e., an earlier T^*_R).
4. A qualification is needed here. We implicitly assume that higher quality staff are not costlier than low quality staff.

Appendix

Claim: If $\omega_1 = 1$, then the rewriting effort is directly proportional to the complexity of the old system and is given by $\frac{1}{\omega_0} F(M(T_R))$.

Proof: Since $F(M(T_R)) = \omega_0 L^{\omega_1} (T_N - T_R)$, therefore $L = \left(\frac{F(M(T_R))}{\omega_0 (T_N - T_R)} \right)^{\frac{1}{\omega_1}}$ and the rewriting effort $L(T_N - T_R) = \left(\frac{F(M(T_R))}{\omega_0} \right)^{\frac{1}{\omega_1}} (T_N - T_R)^{1 - \frac{1}{\omega_1}}$. Thus, when $\omega_1 = 1$, $L(T_N - T_R) = \frac{1}{\omega_0} F(M(T_R))$.

Lemma 1: Let the functions $f_1(t), f_2(t), g(t)$ be defined in the interval $0 \leq t \leq T$ and

1. $g(t_2) \geq g(t_1)$ for $0 \leq t_1 \leq t_2 \leq T$;
2. $f_1(t_2) \leq f_1(t_1)$ for $0 \leq t_1 \leq t_2 \leq T$;
3. $f_2(t_2) \leq f_2(t_1)$ for $0 \leq t_1 \leq t_2 \leq T$;
4. $f_1(t) < f_2(t)$ for $0 \leq t \leq T$.

If $f_1(T_1) = g(T_1)$ and $f_2(T_2) = g(T_2)$, then $T_1 < T_2$.

Proof: By contradiction.

Assume $T_2 \leq T_1$, i.e. $T_2 = T_1 - \delta$ for some $\delta \geq 0$. Then $f_2(T_2) = f_2(T_1 - \delta) = g(T_1 - \delta) \leq g(T_1) = f_1(T_1)$. Therefore, we have $f_1(T_1) \geq f_2(T_1 - \delta) > f_1(T_1 - \delta)$ which contradicts the monotonic non-increasing property of $f_1(t)$. Therefore the assumption $T_2 \leq T_1$ is not valid and $T_1 < T_2$.

Lemma 2: Let the functions $g_1(t), g_2(t), f(t)$ be defined in the interval $0 \leq t \leq T$ and

1. $f(t_2) \leq f(t_1)$ for $0 \leq t_1 \leq t_2 \leq T$;
2. $g_1(t_2) \geq g_1(t_1)$ for $0 \leq t_1 \leq t_2 \leq T$;
3. $g_2(t_2) \geq g_2(t_1)$ for $0 \leq t_1 \leq t_2 \leq T$;
4. $g_1(t) < g_2(t)$ for $0 \leq t \leq T$.

If $g_1(T_1) = f(T_1)$ and $g_2(T_2) = f(T_2)$, then $T_1 > T_2$.

Proof: By contradiction.

Assume $T_2 \geq T_1$, i.e. $T_2 = T_1 + \delta$ for some $\delta \geq 0$. Then $g_2(T_2) = g_2(T_1 + \delta) = f(T_1 + \delta) \leq f(T_1) = g_1(T_1)$. Therefore, we have $g_1(T_1) \geq g_2(T_1 + \delta) > g_1(T_1 + \delta)$ which contradicts the monotonic non-decreasing property of $g_1(t)$. Therefore the assumption $T_2 \geq T_1$ is not valid and $T_1 > T_2$.

Proposition 1. If $\delta < \frac{\ln(\beta_M + 1)}{\beta_M}$, solution to [G1] exists and is unique. At optimal $T_R^* = T_N^*$ and T_R^* is characterized by:

$$\alpha_M^n \left[e^{\delta T_R^*} M(T - T_R^*) \right]^{\beta_M^n} = \alpha_M^o M(T_R^*)^{\beta_M^o} + \frac{\theta_m}{\omega_0}. \quad (5.1)$$

Proof. The cumulative effort of maintenance can be easily shown to be:

$$E(T_R, T_N) = \frac{\alpha_M^o \cdot M(T_N)^{\beta_M^o + 1}}{\beta_M^o + 1} + \left[\frac{\alpha_M^n (e^{\delta T_R} \cdot M(T - T_R))^{\beta_M^n + 1}}{\beta_M^n + 1} \right] + \frac{1}{\omega_0} \cdot (\theta_0 + \theta_m M(T_R)).$$

Since

$$\frac{\partial^2 E(T_N, T_R)}{\partial T_N^2} = \alpha_M^o \beta_M^o M(T_N)^{\beta_M^o - 1} m(0)^2 e^{2\delta T_N} + \alpha_M^o M(T_N)^{\beta_M^o} m(0) \delta > 0,$$

$$\frac{\partial^2 E(T_N, T_R)}{\partial T_R^2} = \alpha_M^n m(0)^2 e^{\delta T_R} [e^{\delta T_R} M(T - T_R)]^{\beta_M^n - 1} \cdot [(\beta_M^n + 1) e^{\delta T_R} - e^{\delta T}],$$

and

$$\frac{\partial^2 E(T_N, T_R)}{\partial T_N \partial T_R} = 0.$$

If $\delta < \frac{\ln(\beta_M^n + 1)}{\beta_M^n}$ then $\delta < \frac{\ln(\beta_M^n + 1)}{T - T_R}$, or $\delta T < \ln(\beta_M^n + 1) + \delta T_R$, i.e. $e^{\delta T} < (\beta_M^n + 1)e^{\delta T_R}$, which implies that $\frac{\partial^2 E(T_N, T_R)}{\partial T_R^2} > 0$. Therefore the function $E(T_N, T_R)$ is convex in T_N and T_R (see Chan, Chung and Ho 1994, for the proof of a more realistic scenario).

Let γ be a Lagrangean multiplier and $L(E(T_R, T_N), \gamma) = E(T_R, T_N) + \gamma(T_R - T_N)$ be the Lagrangean function, if $E(T_N, T_R)$ is convex in T_R and T_N , the solution to [G1] exists and is unique. Let $T_R^* \in (0, T)$ and T_N^* be the unique solution to [G1]. The Kuhn-Tucker Conditions are

$$\alpha_M^o \cdot M(T_N^*)^{\beta_M^o} (m(0)e^{\delta T_N^*}) - \gamma = 0, \quad (5.2)$$

$$-\alpha_M^n [e^{\delta T_R^*} \cdot M(T - T_R^*)]^{\beta_M^n} m(0)e^{\delta T_R^*} + \frac{\theta_m}{\omega_0} m(0)e^{\delta T_R^*} + \gamma = 0, \quad (5.3)$$

$$\gamma(T_R^* - T_N^*) = 0, \quad (5.4)$$

$$\gamma \geq 0. \quad (5.5)$$

Since $\alpha_M^o \cdot M(T_N^*)^{\beta_M^o} (m(0)e^{\delta T_N^*}) > 0$, therefore, from (5.2), $\gamma > 0$ and from (5.4), $T_R^* = T_N^*$. Rearranging terms in (5.2) and (5.3), we have

$$\alpha_M^o \cdot M(T_R^*)^{\beta_M^o} (m(0)e^{\delta T_R^*}) = \gamma \quad (5.6)$$

and

$$\alpha_M^n [e^{\delta T_R^*} \cdot M(T - T_R^*)]^{\beta_M^n} m(0)e^{\delta T_R^*} = \frac{\theta_m}{\omega_0} m(0)e^{\delta T_R^*} + \gamma, \quad (5.7)$$

Substituting (5.6) into (5.7) and simplifying, we obtain the required optimality condition.

Implication 1.1: T_R^* increases with δ when $T_R^* < \frac{T}{\sqrt{2}}$ and decreases with δ when $T_R^* > \frac{T}{\sqrt{2}}$.

Proof: Differentiate the optimality condition given in Proposition 1 with respect to δ , we have

$$\begin{aligned} & \alpha_M^n \beta_M^n \left[\frac{m(0)}{\delta} (e^{\delta T} - e^{\delta T_R^*}) \right]^{\beta_M^n - 1} \left[-\frac{m(0)}{\delta^2} (e^{\delta T} - e^{\delta T_R^*}) + \frac{m(0)}{\delta} (T e^{\delta T} - T_R^* e^{\delta T_R^*}) - m(0) e^{\delta T_R^*} \frac{dT_R^*}{d\delta} \right] \\ &= \alpha_M^o \beta_M^o \left[\frac{m(0)}{\delta} (e^{\delta T_R^*} - 1) \right]^{\beta_M^o - 1} \left[-\frac{m(0)}{\delta^2} (e^{\delta T_R^*} - 1) + \frac{m(0)}{\delta} (T_R^* e^{\delta T_R^*}) + m(0) e^{\delta T_R^*} \frac{dT_R^*}{d\delta} \right]. \end{aligned}$$

Rearranging terms and simplifying, we obtain

$$\begin{aligned} & \left\{ \alpha_M^n \beta_M^n \left[\frac{m(0)}{\delta} (e^{\delta T} - e^{\delta T_R^*}) \right]^{\beta_M^n - 1} + \alpha_M^o \beta_M^o \left[\frac{m(0)}{\delta} (e^{\delta T_R^*} - 1) \right]^{\beta_M^o - 1} \right\} m(0) e^{\delta T_R^*} \frac{dT_R^*}{d\delta} \\ &= \alpha_M^n \beta_M^n \left[\frac{m(0)}{\delta} (e^{\delta T} - e^{\delta T_R^*}) \right]^{\beta_M^n - 1} \frac{m(0)}{\delta} \left[(T e^{\delta T} - T_R^* e^{\delta T_R^*}) - \frac{1}{\delta} (e^{\delta T} - e^{\delta T_R^*}) \right] \\ & \quad + \alpha_M^o \beta_M^o \left[\frac{m(0)}{\delta} (e^{\delta T_R^*} - 1) \right]^{\beta_M^o - 1} \frac{m(0)}{\delta} \left[\frac{1}{\delta} (e^{\delta T_R^*} - 1) - T_R^* e^{\delta T_R^*} \right]. \end{aligned}$$

¹An extensive numerical simulation experiment shows that T_R^* is less than $\frac{T}{\sqrt{2}}$ when the new technology is superior than the old technology.

Let $M = \min \left\{ \alpha_M^n \beta_M^n \left[\frac{m(0)}{\delta} (e^{\delta T} - e^{\delta T_R^*}) \right]^{\beta_M^n - 1} \frac{m(0)}{\delta}, \alpha_M^o \beta_M^o \left[\frac{m(0)}{\delta} (e^{\delta T_R^*} - 1) \right]^{\beta_M^o - 1} \frac{m(0)}{\delta} \right\}$. Since both terms are positive, therefore $M > 0$ and

$$\begin{aligned} & \left\{ \alpha_M^n \beta_M^n \left[\frac{m(0)}{\delta} (e^{\delta T} - e^{\delta T_R^*}) \right]^{\beta_M^n - 1} + \alpha_M^o \beta_M^o \left[\frac{m(0)}{\delta} (e^{\delta T_R^*} - 1) \right]^{\beta_M^o - 1} \right\} m(0) e^{\delta T_R^*} \frac{dT_R^*}{d\delta} \\ & > M \left[(T e^{\delta T} - T_R^* e^{\delta T_R^*}) - \frac{1}{\delta} (e^{\delta T} - e^{\delta T_R^*}) + \frac{1}{\delta} (e^{\delta T_R^*} - 1) - T_R^* e^{\delta T_R^*} \right]. \end{aligned}$$

Using linear approximations for $e^{\delta T}$ and $e^{\delta T_R^*}$ when δT and δT_R^* are small, we have

$$\begin{aligned} & \left\{ \alpha_M^n \beta_M^n \left[\frac{m(0)}{\delta} (e^{\delta T} - e^{\delta T_R^*}) \right]^{\beta_M^n - 1} + \alpha_M^o \beta_M^o \left[\frac{m(0)}{\delta} (e^{\delta T_R^*} - 1) \right]^{\beta_M^o - 1} \right\} m(0) e^{\delta T_R^*} \frac{dT_R^*}{d\delta} \\ & > M [T(1 + \delta T) - T_R^*(1 + \delta T_R^*) - (T - T_R^*) + T_R^* - T_R^*(1 + \delta T_R^*)] = M\delta [T^2 - 2T_R^{*2}]. \end{aligned}$$

Therefore

$$\frac{dT_R^*}{d\delta} > 0 \text{ for } T_R^* < \frac{T}{\sqrt{2}} \quad \text{and} \quad \frac{dT_R^*}{d\delta} < 0 \text{ for } T_R^* > \frac{T}{\sqrt{2}}.$$

Implication 1.2: T_R^* is independent of θ_0 .

Proof: It follows directly because the expression for T_R^* does not involve θ_0 .

Implication 1.3: T_R^* decreases with θ_m .

Proof: Let $f(T_R) = \alpha_M^n \left[\frac{m(0)}{\delta} (e^{\delta T} - e^{\delta T_R}) \right]^{\beta_M^n}$, $g_1(T_R) = \alpha_M^o M(T_R)^{\beta_M^o} + \frac{\theta_{m1}}{\omega_0}$, and $g_2(T_R) = \alpha_M^o M(T_R)^{\beta_M^o} + \frac{\theta_{m2}}{\omega_0}$ be defined on the interval $[0, T]$ with $0 < \theta_{m1} < \theta_{m2}$ and $g_1(T_R^*) = f(T_R^*)$, $g_2(T_R^{**}) = f(T_R^{**})$. It can be seen that $f(T_R)$ is monotonic non-increasing and both $g_1(T_R)$ and $g_2(T_R)$ are monotonic non-decreasing in the interval $[0, T]$. For all T_R in the interval $(0, T)$, $g_2(T_R) - g_1(T_R) = \frac{1}{\omega_0}(\theta_{m2} - \theta_{m1}) > 0$ for $\omega_0 > 0$. By Lemma 2, $T_R^{**} < T_R^*$.

Implication 1.4: T_R^* increases with ω_0 .

Proof: Let $f(T_R) = \alpha_M^n \left[\frac{m(0)}{\delta} (e^{\delta T} - e^{\delta T_R}) \right]^{\beta_M^n}$, $g_1(T_R) = \alpha_M^o M(T_R)^{\beta_M^o} + \frac{\theta_{m1}}{\omega_0}$, and $g_2(T_R) = \alpha_M^o M(T_R)^{\beta_M^o} + \frac{\theta_{m2}}{\omega_0}$ be defined on the interval $[0, T]$ with $\omega_{02} > \omega_{01} > 0$ and $g_1(T_R^*) = f(T_R^*)$, $g_2(T_R^{**}) = f(T_R^{**})$. Since $f(T_R)$ is monotonic non-increasing and $g_1(T_R), g_2(T_R)$ are monotonic non-decreasing in $(0, T)$ and $g_2(T_R) - g_1(T_R) = \theta_m \left(\frac{1}{\omega_{02}} - \frac{1}{\omega_{01}} \right) < 0$, by Lemma 2, we have $T_R^{**} > T_R^*$.

Implication 1.5: T_R^* decreases with a decrease in α_M^n .

Proof: Let $g(T_R) = \alpha_M^o M(T_R)^{\beta_M^o} + \frac{\theta_m}{\omega_0}$, $f_1(T_R) = \alpha_M^{n1} \left[\frac{m(0)}{\delta} (e^{\delta T} - e^{\delta T_R}) \right]^{\beta_M^{n1}}$, and $f_2(T_R) = \alpha_M^{n2} \left[\frac{m(0)}{\delta} (e^{\delta T} - e^{\delta T_R}) \right]^{\beta_M^{n2}}$ be defined on the interval $[0, T]$ with $0 < \alpha_M^{n2} < \alpha_M^{n1}$. If $f_1(T_R^*) = g(T_R^*)$ and $f_2(T_R^{**}) = g(T_R^{**})$, then $g(T_R)$ is monotonic non-decreasing, $f_1(T_R)$ and $f_2(T_R)$ are monotonic non-increasing in $(0, T)$. $f_2(T_R) - f_1(T_R) = (\alpha_M^{n2} - \alpha_M^{n1}) \left[\frac{m(0)}{\delta} (e^{\delta T} - e^{\delta T_R}) \right]^{\beta_M^{n1}} < 0$. By Lemma 1, $T_R^{**} < T_R^*$.

Implication 1.6: T_R^* decreases with a decrease in β_M^n .

Proof: Let $g(T_R) = \alpha_M^o M(T_R)^{\beta_M^o} + \frac{\theta_m}{\omega_0}$, $f_1(T_R) = \alpha_M^{n1} \left[\frac{m(0)}{\delta} (e^{\delta T} - e^{\delta T_R}) \right]^{\beta_M^{n1}}$, and $f_2(T_R) = \alpha_M^n \left[\frac{m(0)}{\delta} (e^{\delta T} - e^{\delta T_R}) \right]^{\beta_M^n}$ be defined on the interval $[0, T]$ with $0 < \beta_M^{n2} < \beta_M^{n1}$. If $f_1(T_R^*) = g(T_R^*)$ and $f_2(T_R^{**}) = g(T_R^{**})$. Note that $g(T_R)$ is monotonic non-decreasing, $f_1(T_R)$ and $f_2(T_R)$ are monotonic non-increasing in $(0, T)$. In addition, $f_2(T_R) - f_1(T_R) = \alpha_M^n \left\{ \left[\frac{m(0)}{\delta} (e^{\delta T} - e^{\delta T_R}) \right]^{\beta_M^{n2}} - \left[\frac{m(0)}{\delta} (e^{\delta T} - e^{\delta T_R}) \right]^{\beta_M^{n1}} \right\} < 0$. By Lemma 1, $T_R^{**} < T_R^*$.

Proposition 2. If $F_M^n < \frac{\theta_m}{\omega_0}$ and $\delta < \frac{\ln 2}{T}$, solution to [G1] exists and is unique. At optimal, $T_R^* = T_N^*$ and T_R^* is characterized by:

$$\frac{m(0)V_M^n}{\delta}(e^{\delta T} - e^{\delta T_R^*}) + F_M^n = \frac{m(0)V_M^o}{\delta}(e^{\delta T_R^*} - 1) + F_M^o + \frac{\theta_m}{\omega_0}. \quad (5.8)$$

Proof. It can be easily shown that cumulative maintenance effort is given by

$$E(T_R, T_N) = \left(\frac{m(0)F_M^o}{\delta} - \frac{m(0)^2V_M^o}{\delta^2} \right) (e^{\delta T_N} - 1) + \frac{m(0)^2V_M^o}{2\delta^2} (e^{2\delta T_N} - 1) \\ + \frac{m(0)F_M^n}{\delta} (e^{\delta T} - e^{\delta T_R}) + \frac{m(0)^2V_M^n}{2\delta^2} (e^{\delta T} - e^{\delta T_R})^2 + \frac{1}{\omega_0} \left(\theta_0 + \frac{m(0)\theta_m}{\delta} e^{\delta T_R} - \frac{m(0)\theta_0}{\delta} \right).$$

Since

$$\frac{\partial^2 E(T_N, T_R)}{\partial T_N^2} = m(0)F_M^o \delta e^{\delta T_N} + m(0)^2V_M^o (2e^{2\delta T_N} - e^{\delta T_N}) > 0,$$

$$\frac{\partial^2 E(T_N, T_R)}{\partial T_R^2} = m(0)\delta e^{\delta T_R} \left(\frac{\theta_m}{\omega_0} - F_M^n \right) + m(0)^2V_M^n [e^{\delta T_R} (2e^{\delta T_R} - e^{\delta T})],$$

and

$$\frac{\partial^2 E(T_N, T_R)}{\partial T_N \partial T_R} = 0.$$

If $F_M^n < \frac{\theta_m}{\omega_0}$ then $(\frac{\theta_m}{\omega_0} - F_M^n) > 0$. Also if $\delta < \frac{\ln 2}{T}$, then $\delta < \frac{\ln 2}{T - T_R}$ or $e^{\delta T} < 2e^{\delta T_R}$. Therefore $\frac{\partial^2 E(T_N, T_R)}{\partial T_R^2} > 0$. Hence, $E(T_N, T_R)$ is convex in T_N and T_R (see Chan, Chung and Ho 1994, for the proof of a more realistic senario).

Let $L(E(T_R, T_N)) = E(T_R, T_N) + \gamma(T_R - T_N)$ be the Lagrangean. Provided $E(T_R, T_N)$ is convex in T_R and T_N , the optimal solution to [G1] exists and is unique. Let $T_R^* \in (0, T)$ and T_N^* be the unique solution to [G1]. The K-T conditions are:

$$\left(\frac{m(0)F_M^o}{\delta} - \frac{m(0)^2V_M^o}{\delta^2} \right) \delta e^{\delta T_N^*} + \frac{m(0)^2V_M^o}{2\delta^2} 2\delta e^{2\delta T_N^*} - \gamma = 0, \quad (5.9)$$

$$\frac{m(0)F_M^n}{\delta} (-\delta e^{\delta T_R^*}) + \frac{m(0)^2V_M^n}{2\delta^2} \cdot 2(e^{\delta T} - e^{\delta T_R^*})(-\delta e^{\delta T_R^*}) + \frac{m(0)\theta_m}{\omega_0} e^{\delta T_R^*} + \gamma = 0, \quad (5.10)$$

$$\gamma(T_R^* - T_N^*) = 0, \quad (5.11)$$

$$\gamma > 0. \quad (5.12)$$

From (5.9), we have:

$$m(0)F_M^o e^{\delta T_N^*} + \frac{m(0)^2V_M^o}{2\delta} (e^{2\delta T_N^*} - e^{\delta T_N^*}) = \gamma > 0. \quad (5.13)$$

From (5.11), $T_R^* = T_N^*$. Rearranging (5.10) gives

$$m(0)F_M^n e^{\delta T_R^*} + \frac{m(0)V_M^n}{\delta} (e^{\delta T} - e^{\delta T_R^*}) e^{\delta T_R^*} = \frac{m(0)\theta_m}{\omega_0} e^{\delta T_R^*} + \gamma. \quad (5.14)$$

Substituting (5.13) into (5.14) and simplifying, we obtain the required optimality condition.

The managerial insights with respect to $\delta, \theta_0, \theta_m, \omega_0$ can be proven in a similar way. The condition for T_R^* increases with δ however changes somewhat. The new condition is: if $r = \frac{V_M^o}{V_M^n}$, then T_R^* increases with δ when $T_R^* < \frac{T}{\sqrt{1+r}}$ and T_R^* decreases with δ when $T_R^* > \frac{T}{\sqrt{1+r}}$.

Implication 2.1: T_R^* decrease with F_M^o and V_M^o .

Proof: Let $f(T_R) = \frac{m(0)V_M^o}{\delta}(e^{\delta T} - e^{\delta T_R}) + F_M^n$, $g_1(T_R) = \frac{m(0)V_M^{o1}}{\delta}(e^{\delta T_R} - 1) + F_M^{o1} + \frac{\theta_m}{\omega_0}$, and $g_2(T_R) = \frac{m(0)V_M^{o2}}{\delta}(e^{\delta T_R} - 1) + F_M^{o2} + \frac{\theta_m}{\omega_0}$ be defined on the interval $[0, T]$ with $F_M^{o2} > F_M^{o1} > 0$ and $V_M^{o2} > V_M^{o1} > 0$. If $g_1(T_R^*) = f(T_R^*)$, $g_2(T_R^{**}) = f(T_R^{**})$. Since $f(T_R)$ is monotonic non-increasing and $g_1(T_R), g_2(T_R)$ are monotonic non-decreasing in $(0, T)$ and $g_2(T_R) - g_1(T_R) = \frac{m(0)}{\delta}(e^{\delta T_R} - 1)(V_M^{o2} - V_M^{o1}) + (F_M^{o2} - F_M^{o1}) > 0$. Therefore, using Lemma 2, $T_R^{**} < T_R^*$.

Implication 2.2: T_R^* increases with increases in F_M^n and V_M^n .

Proof: Let $g(T_R) = \frac{m(0)V_M^o}{\delta}(e^{\delta T_R} - 1) + F_M^o + \frac{\theta_m}{\omega_0}$, $f_1(T_R) = \frac{m(0)V_M^{n1}}{\delta}(e^{\delta T} - e^{\delta T_R}) + F_M^{n1}$, and $f_2(T_R) = \frac{m(0)V_M^{n2}}{\delta}(e^{\delta T} - e^{\delta T_R}) + F_M^{n2}$ be defined on the interval $[0, T]$ with $F_M^{n2} > F_M^{n1}$ and $V_M^{n2} > V_M^{n1}$. If $f_1(T_R^*) = g(T_R^*)$, $f_2(T_R^{**}) = g(T_R^{**})$. Since $g(T_R)$ is monotonic non-decreasing and $f_1(T_R), f_2(T_R)$ are monotonic non-increasing in $(0, T)$ and $f_2(T_R) - f_1(T_R) = \frac{m(0)}{\delta}(e^{\delta T} - e^{\delta T_R})(V_M^{n2} - V_M^{n1}) + (F_M^{n2} - F_M^{n1}) > 0$. Therefore, using Lemma 2, $T_R^{**} > T_R^*$.

Proposition 3: If $\delta < \frac{\ln(\beta_M + 1)}{T}$, solution to [G2] exists and is unique. At optimal, T_R^* and T_N^* are characterized by:

$$\alpha_M^n \left[\frac{m(0)}{\delta} (e^{\delta T} - e^{\delta T_R^*}) \right]^{\beta_M^n} = \alpha_M^o M(T_N^*)^{\beta_M^o} e^{\delta(T_N^* - T_R^*)} \left(1 + \frac{\theta_m}{\omega_0 L_R^{\omega_1}} m(T_R^*) \right) + \frac{\theta_m}{\omega_0} L_R^{1-\omega_1}.$$

Proof: It can be easily shown that the total cumulative effort is given by:

$$E(T_R, T_N) = \frac{\alpha_M^o \cdot M(T_N)^{\beta_M^o + 1}}{\beta_M^o + 1} + \frac{\alpha_M^n (e^{\delta T_R} \cdot M(T - T_R))^{\beta_M^n + 1}}{\beta_M^n + 1} + \frac{1}{\omega_0} (\theta_0 + \theta_m M(T_R)) \cdot L_R^{1-\omega_1}.$$

Similar to Proposition 1, we can show that if $\delta < \frac{\ln(\beta_M + 1)}{T}$, $E(T_N, T_R)$ is also convex in T_N and T_R . Substitute $T_N = T_R + \frac{F(M(T_R))}{\omega_0 L_R^{\omega_1}}$ into $E(T_R, T_N)$. Using first-order condition for optimality,

$$\begin{aligned} \frac{dE}{dT_R} &= \alpha_M^o M(T_N^*)^{\beta_M^o} m(T_N^*) \frac{dT_N^*}{dT_R^*} - \alpha_M^n (e^{\delta T_R} \cdot M(T - T_R^*))^{\beta_M^n} m(0) e^{\delta T_R^*} \\ &+ \frac{\theta_m}{\omega_0} L_R^{1-\omega_1} \cdot m(T_R^*) = 0. \end{aligned}$$

After rearrangement and simplification, we obtain the required optimality condition.

Proposition 4. If $F_M^n < \frac{L_R^{1-\omega_1} \theta_m}{L_R^{\omega_1}}$ and $\delta < \frac{\ln 2}{T}$, solution to [G2] exists and is unique. At optimal, T_R^* and T_N^* are uniquely determined by:

$$F_M^n + \frac{m(0)V_M^n}{\delta} (e^{\delta T} - e^{\delta T_R^*}) = \left[F_M^o + \frac{m(0)V_M^o}{\delta} (e^{\delta T_N^*} - 1) \right] e^{\delta(T_N^* - T_R^*)} \left[1 + \frac{\theta_m}{\omega_0 L_R^{\omega_1}} m(T_R^*) \right] + \frac{L_R^{1-\omega_1} \theta_m}{\omega_0} (e^{\delta T_N^*} - 1) \quad (5.15)$$

Proof. **Proof:** It can be shown easily that the total cumulative effort is given by

$$\begin{aligned} E(T_R, T_N) &= \left(\frac{m(0)F_M^o}{\delta} - \frac{m(0)^2 V_M^o}{\delta^2} \right) (e^{\delta T_N} - 1) + \frac{m(0)^2 V_M^o}{2\delta^2} (e^{2\delta T_N} - 1) \\ &+ \frac{m(0)F_M^n}{\delta} (e^{\delta T} - e^{\delta T_R}) + \frac{m(0)^2 V_M^n}{2\delta^2} (e^{\delta T} - e^{\delta T_R})^2 \\ &+ \frac{L_R^{1-\omega_1}}{\omega_0} \left(\theta_0 + \frac{m(0)\theta_m}{\delta} e^{\delta T_R} - \frac{m(0)\theta_m}{\delta_m} \right). \end{aligned}$$

Similar to Proposition 2, if $F_M^n < \frac{L_r^{1-\omega_1} \theta_m}{\omega_0}$ and $\delta < \frac{\ln 2}{T}$, we can show that $E(T_N, T_R)$ is convex in T_N and T_R . The first-order optimality condition with respect to T_R^* is given by

$$\begin{aligned} \frac{dE}{dT_R} = & \left(\frac{m(0)F_M^o}{\delta} - \frac{m(0)^2 V_M^o}{\delta^2} \right) e^{\delta T_N^*} \cdot \delta \cdot \frac{dT_N^*}{dT_R^*} + \frac{m(0)^2 V_M^o}{2\delta^2} e^{2\delta T_N^*} \cdot 2\delta \frac{dT_N^*}{dT_R^*} \\ & + \frac{m(0)F_M^n}{\delta} (-\delta e^{\delta T_R^*}) + \frac{m(0)^2 V_M^n}{2\delta^2} \cdot 2(e^{\delta T} - e^{\delta T_R^*})(-\delta e^{\delta T_R^*}) + \frac{L_R^{1-\omega_1}}{\omega_0} \left(\frac{m(0)\theta_m}{\delta} \cdot \delta e^{\delta T_R^*} \right) = 0. \end{aligned}$$

Substituting $\frac{dT_N^*}{dT_R^*} = \left(1 + \frac{\theta_m}{\omega_0 L_R^{\omega_1}} m(T_R^*)\right)$ and simplifying gives the required optimality condition.

Implication 3.1: T_R^* increases with ω_1 .

Proof: a) Log-Linear case: Let $f(T_R^*) = \alpha_M^n \left[\frac{m(0)}{\delta} (e^{\delta T} - e^{\delta T_R^*}) \right]^{\beta_M^n}$ and $g(T_R^*) = \alpha_M^o M(T_N^*)^{\beta_M^o} e^{\delta \left(\frac{F(M(T_R^*))}{\omega_0 L_R^{\omega_1}} \right)} \left(1 + \frac{\theta_m}{\omega_0 L_R^{\omega_1}} m(T_R^*)\right) + \frac{\theta_m}{\omega_0} L_R^{1-\omega_1}$. Since $f(T_R^*)$ is monotonic non-increasing and $g(T_R^*)$ is monotonic non-decreasing in $[0, T]$ and if ω_1 increases, $g(T_R^*)$ decreases and by Lemma 2, T_R^* increases. b) Linear case: Can be proven similarly.

Implication 3.2: T_R^* decreases with θ_0 .

Proof: a) Log-linear case: Let $f(T_R^*) = \alpha_M^n \left[\frac{m(0)}{\delta} (e^{\delta T} - e^{\delta T_R^*}) \right]^{\beta_M^n}$ and $g(T_R^*) = \alpha_M^o M(T_N^*)^{\beta_M^o} e^{\delta \left(\frac{\theta_0 + \theta_m M(T_R^*)}{\omega_0 L_R^{\omega_1}} \right)} \left(1 + \frac{\theta_m}{\omega_0 L_R^{\omega_1}} m(T_R^*)\right) + \frac{\theta_m}{\omega_0} L_R^{1-\omega_1}$. Since $f(T_R^*)$ is monotonic non-increasing and $g(T_R^*)$ is monotonic non-decreasing in $[0, T]$ and if θ_0 increases, $g(T_R^*)$ increases and by Lemma 2, T_R^* decreases. b) Linear case: Can be proven similarly.

Implication 3.3: $T_N^* - T_R^*$ decreases with a decrease in α_M^n , F_M^n , and V_M^n .

Proof: Like before, it can be easily proven that T_R^* decreases with a decrease in α_M^n , F_M^n , and V_M^n . Since $T_N^* - T_R^* = \frac{\theta_0 + \theta_m M(T_R^*)}{\omega_0 L_R^{\omega_1}}$, we have the required implication.

Implication 3.4: $T_N^* - T_R^*$ decreases with a decrease in β_M^n .

Proof: Like before, it can be easily proven that T_R^* decreases with a decrease in β_M^n . Since $T_N^* - T_R^* = \frac{\theta_0 + \theta_m M(T_R^*)}{\omega_0 L_R^{\omega_1}}$, we have the required implication.