Scheduling Deteriorating Jobs on a Single Machine Subject to Breakdowns*

Xiaoqiang CAI[1], Xianyi WU[2], and Xian ZHOU[3]
[1]Department of Systems Engineering and Engineering Management
The Chinese University of Hong Kong, Hong Kong
[2]Department of Statistics and Actuarial Science
East China Normal University, Shanghai
[3]Department of Actuarial Studies
Macquarie University, Australia

Abstract

We investigate the problem of scheduling a set of jobs to minimize the expected makespan or the variance of the makespan. The jobs are subject to deteriorations, which are expressed as linear increments of the processing requirements. The machine is subject to preemptive-resume breakdowns with exponentially distributed uptimes and downtimes. It has been well known in the classical models that the expectation and variance of the makespan of deteriorating jobs can be minimized analytically by an index policy if no machine breakdowns are involved. Such basic features, however, change dramatically when breakdowns and deteriorations are present together. In this paper, we derive the conditions for jobs to be *processible* in the sense that they will be eventually completed, and the characteristics of the time that a job occupies the machine. We further find that the expected makespan can still be minimized by a simple index policy that is independent of the breakdown process, but this is no longer the case for the variance of the makespan.

Key words and phrases: Stochastic scheduling; machine breakdowns; preemptive-resume; deteriorations, processibility of job.

*This research is partially supported by the Research Grants Council of Hong Kong under Earmarked Grant No. CUHK4170/03E and Direct Grant No. 2050415, NSFC/RGC Joint Research Grant No. N-CUHK442/05 & No. 70671043, and Macquarie Safety Net Scheme.

1 Introduction

In the mainstream of theoretical researches and practical applications of scheduling, one of the broadly adopted assumptions is that the processing time of each job is time-invariant. More precisely, no matter when a job is selected to start its processing, its processing time is always independent of the start time of the operation, see for example Pinedo (2002). In practice, however, there are many situations where the processing time is time-varying in the sense that it may be a function of the time at which it starts its processing (referred to as *start-time-dependent scheduling*). While the time-invariant assumption may reflect (or approximate) the real life in some cases, it is hardly applicable in general, and is often an over-simplified picture of the reality so as to take the advantage of computational convenience. Many practical instances have been reported that severely violate the time-invariant assumption, and significant progress regarding the time-dependent scheduling has been made in the literature. Applications of these models have been found in, among others, fire fighting, financial management, food processing, maintenance, resource allocation, military objective searching, national defense, and computer science. In all these real examples, the delay of the start time of processing will impact (increase or decrease) the overall efforts (time, cost, etc.) to accomplish the task. Models for machine scheduling problems with start-time-dependent processing times have attracted increasing attentions in the past two decades and many remarkable findings have been reported in this area. Under an important group of models, job processing times are assumed to be nondecreasing in the start time, which is referred as *deterioration* (as waiting would increase a job's processing time). For more detailed discussions of the literature and the research history in this field, refer to the two survey papers by Cheng, Ding and Lin (2004) and Alidaee and Womer (1999) and the references therein.

Another assumption that is usually taken for granted in classical scheduling models is that the machine is always ready to attend a job. In reality, however, there are many circumstances where the machine is subject to stochastic breakdowns from time to time, which results in interruptions of job processing. Machine breakdowns have long been recognized as a common phenomenon in many real situations such as computing, manufacturing, refinery, and so on. In particular, a machine is more likely to be subject to breakdowns when a job takes a long time span to process. Scheduling models considering machine breakdowns have been extensively studied over two decades. By and large, the main research efforts can be broken into two categories. The first is referred to as the preemptive-resume model, including Birge et al. (1990), Cai and Zhou (1999, 2000), Glazebrook (1984, 1987), Li, Braun and Zhao (1998), Mittenthal and Raghavachari (1993), Pinedo and Rammouz (1988), Qi, Yin and Birge (2000a, 2000b), to name just a few. Under this model, when the machine resumes its operation after a breakdown is fixed, the processing of the interrupted job will continue from the point where it was interrupted. The other is generally called the preemptive-repeat model, which assumes that the work done on an unfinished job is totally lost after a machine breakdown and it must be processed from scratch again after the machine is fixed. Significant contributions on this model include Birge et al. (1990), Cai, Sun and Zhou (2003, 2004), Cai, Wu and Zhou (2005, 2009), Frostig (1991), and Glazebrook (1991), among others.

A significant feature of machine breakdown is its companion fixing time (or called down-time in the literature), which has a crucial impact on the processing of a job as well as the information accumulation process. This impact has been shown to usually magnify the difficulty of the scheduling problem significantly. It will be much more serious in the case of start-time-dependent job processing, where any delay of processing due to breakdowns could further increase the processing time/cost. This "double" impact makes the job scheduling much more difficult to tackle when the processing times are start-time-dependent; the worst case is an *unprocessible* job in the sense that its completion time may be infinite with a positive probability.

Examples of facing both increasing processing times and stochastic machine breakdowns occur in many practical settings. For instance, in scheduling steel rolling, the rolling machine may go out of service randomly from time to time, and the temperature of an ingot drops during waiting so that it needs to be re-heated to the temperature required for rolling, leading to increased rolling time. Another example arises in searching for a military object such as a submarine or a group of enemy soldiers under worsening weather conditions or growing darkness, where the searching instrument such as an antisubmarine facility or a scout may breakdown and the searching time may increase due to worsening conditions. Furthermore, in fire fighting, the time required to control a fire would increase if there was a delay in the start of the fire fighting effort, and any failure of facilities results in a breakdown of the fire fighting. In all of these situations, accomplishing a task (job) may be affected by time factor and machine breakdowns. Under such circumstances, the properties of job processing times and the operation system may change significantly, and the complexity of seeking optimal policies may increase dramatically.

In this paper we intend to explore the features of scheduling problems with a set of n deteriorating jobs on a single machine subject to stochastic breakdowns, and the solutions to minimize some performance measures. We shall focus on the deterioration in the form of linearly increasing processing requirements and preemptive-resume machine breakdowns. The contributions of the paper include the following aspects:

- Reformulation of the mechanism of linear deteriorations to allow machine breakdowns. This is done in Section 2.
- The conditions for a job to be *processible* under deterioration and machine breakdowns, see Sections 3 for details.
- The probabilistic features of the model with exponentially distributed uptimes and downtimes, which are derived by means of Laplace transforms and differential equations in Section 4.
- Optimal policies for minimizing the expected makespan, in Section 5.
- Analytical expression of the variance of makespan, and discussions on its solutions and complexity, which are treated in Section 5 also.

Concluding remarks are discussed in Section 6.

2 Model Formulation

Suppose that a set of n jobs are to be processed on a single machine, which are all available at time zero. The machine can process at most one job at a time. We are concerned only with the *static policies* $\lambda = \{i_1, i_2, \ldots, i_n\}$, which are permutations to decide the order of the jobs to be processed.

First we look at the case without job deterioration and machine breakdowns. In such a standard case, each job i is associated with an *initial processing time* X_i, $i = 1, 2, \ldots, n$, which are assumed to be independent of one another.

When the problem is subject to job deterioration and machine breakdowns, the 'true' processing time may defer from X_i. The formulations of deteriorations and breakdowns are elaborated in what follows. In the following two sections, the subscriptions on job index i are suppressed for ease of notation since we only work with the features of an individual job.

Modelling machine breakdowns: Scheduling models for machine breakdowns are generally categorized into two types according to the effects of breakdowns on the job being processed: the *preemptive-resume* model, in which the accumulated processing achievement is not lost at a breakdown; and the *preemptive-repeat* model, in which a breakdown destroys all the processing previously done on the job, see, e.g., Cai, Wu and Zhou (2005). We focus on the breakdown model described below:

- The breakdown process for a job is characterized by a sequence of nonnegative random pairs $\{Y_k, Z_k\}_{k=1}^{\infty}$, where Y_k and Z_k are the durations of the k-th uptime and downtime respectively, which are independently and identically distributed as a typical representative (Y, Z);
- Y and Z are supposed to be independent of each other with distribution functions $G(x)$ and $H(x)$ respectively. We sometimes further assume that Y_k and Z_k follow exponential distributions with rates λ_Y and λ_Z respectively.
- The breakdown processes are independent over jobs;
- The breakdowns are preemptive-resume.

Modelling job deterioration: In the literature, see, e.g., Cheng, Ding and Lin (2004) and Alidaee & Womer (1999), the deterioration for a scheduling system means that the processing time of a job is start-time-dependent. Further, no matter what mechanisms the deterioration follows, it is commonly assumed that once the processing of a job has started, the deterioration effect ceases. This assumption is adequate in many situations without machine breakdowns, at least approximately. When the machine is subject to breakdowns, however, it does not reflect the deterioration in downtimes. In reality, it is unlikely that a job will stop its deterioration once it is selected to be processed. Therefore, the formulation should reflect the deterioration in the downtime as well. While there have been many models (such as linear and piecewise linear) characterizing the deteriorations, for tractability, we consider only the *linear deterioration* model as follows.

- We begin with the definition of *processing requirement at time* t, which is the processing time of the job required to complete the job under the 'standard' conditions (no deterioration and no machine breakdowns). For a deteriorative job, the *processing requirement,* denoted by $X(t)$, is a stochastic process with $X(0) = X$ being the requirement at time zero (referred to as the *initial requirement*).

- Under the linear deterioration assumption, $X(t + \Delta t) = X(t) + \alpha\Delta t$ if the job is idle (not receiving any processing efforts) during the time interval $(t, t + \Delta t]$, where $0 \leq \alpha < 1$; and $X(t + \Delta t) = X(t) - (1 - \alpha)\Delta t$ if it is processed during the time interval $(t, t+\Delta t]$. Here $\alpha\Delta t$ indicates the increment of processing requirement due to the deterioration and Δt stands for the reduction of the processing requirement thanks to the job processing. The processing on the job at any time is a fight against the deterioration with the effect of reducing the processing requirement at rate $(1 - \alpha)$. This model is referred to as a *restless* deterioration.

- Under the preemptive-resume assumption, there will be no loss of processing achievement due to breakdowns. To be specific, if the machine breaks down at time s with instant processing requirement $X(s)$ and then experiences a downtime Z, the job will be reprocessed again at time $s + Z$ with the new processing requirement $X(s) + \alpha Z$. Figure 1 below shows a typical sample path of the requirement process $X(s)$ for a job from time zero to its completion. The sample path is continuous and piecewise linear.

Remark 2.1 *We here illustrate the connection between the newly formulated restless deterioration and the traditional deterioration. If the job starts its processing at time* s *and then is processed continuously till its completion (i.e., no machine breakdowns), then the requirement process of the job is*

$$X(t) = \begin{cases} X + \alpha t & t \leq s \\ X + s - (1-\alpha)t & t > s \end{cases}. \tag{2.1}$$

Clearly, the processing on a job is terminated at the first time when its processing requirement reaches zero. The instant $T(s)$ *at the completion of the job is given by*

$$T(s) = \inf\{t : X(t) = 0\}.$$

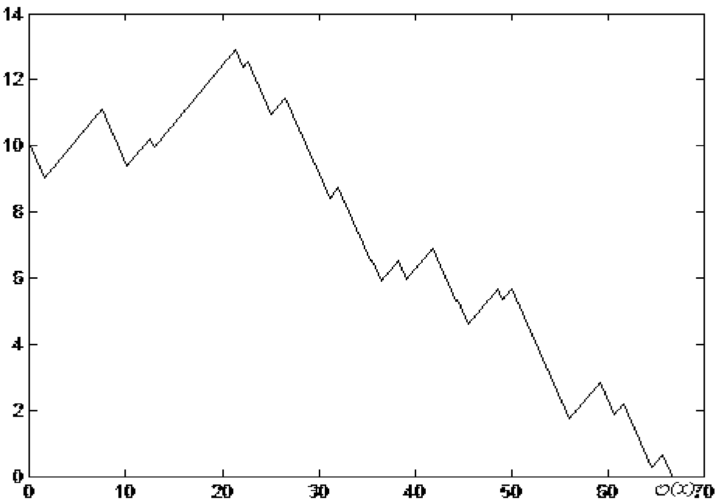

Figure 1: A typical realization of processing requirement

Thus $X(T(s)) = 0$ *and so* $T(s) = (X + s)/(1 - \alpha)$ *by (2.1). Consequently, at time* s*, where the job is selected to be processed, the real processing time (if non-preemptively) is*

$$T(s) - s = \frac{X + s}{1 - \alpha} - s = \frac{X}{1 - \alpha} + \frac{\alpha}{1 - \alpha} s = T(0) + \frac{\alpha}{1 - \alpha} s. \tag{2.2}$$

Here and throughout this paper, we assume $0 < \alpha < 1$*. Equation (2.2) shows that our restless deterioration model coincides with the traditional assumptions, with* $T(0)$ *and* $\alpha/(1 - \alpha)$ *in place of the initial processing time and the deterioration rate in the traditional linear deterioration model.*

Denote $\tau_k = Y_k + Z_k$ to be the k-th processing duration of the job and define a point process $N(t)$ as

$$N(t) = \max\left\{ m : \sum_{k=0}^{m} \tau_k \leq t \right\},$$

which is exactly the frequency of breakdowns by time t, where $\tau_0 = 0$ is used for convenience. Then the processing requirement $X(t)$ can be expressed as

$$X(t) = X + \alpha t - \sum_{k=0}^{N(t)} Y_k - (t - T_{N(t)}) \wedge Y_{N(t)+1}, \tag{2.3}$$

where $T_m = \sum_{k=0}^{m} \tau_k$, X is the initial processing requirement of the job, and $\wedge$ is the minimum operator defined by $a \wedge b = \min(a, b)$. $X(t)$ appears as an intractable stochastic process since $N(t)$ is not independent of $\{Y_k\}$.

One of the most important quantities is the duration for which the job will occupy the machine if it is nonpreemptively processed from start to completion, which is referred to as the *occupying time*. Denote $\mathcal{O}(x)$ to be the occupying time of the job with initial requirement x, if it starts its processing at time zero. $\mathcal{O}(x)$ is the smallest solution to the equation $X(\mathcal{O}(x)) = 0$. Moreover, if the job starts at some time $s > 0$, the occupying time, denoted by $\mathcal{O}(x, s)$ at this point, can be calculated by replacing x with $x + \alpha s$ as $\mathcal{O}(x, s) = \mathcal{O}(x + \alpha s)$, which implies that to get the occupying time for different start time, we only need to calculate $\mathcal{O}(x)$ for all $x > 0$.

Objective function: Let $C_i(\lambda)$ denote the completion time of job i and $C_{(j)}(\lambda)$ the completion time of the j-th processed job under any policy λ. The objective is to find optimal policies that minimize the following expectation (EM) and variance (VM) of makespan :

$$\mathrm{EM}(\lambda) = \mathrm{E}\left[C_{\max}\right], \qquad \mathrm{VM}(\lambda) = \mathrm{Var}\left[C_{\max}\right], \tag{2.4}$$

where $C_{\max} = C_{(n)}$ is the makespan of the n jobs, i.e., the completion time of the job last processed.

Remark 2.2 *Glazebrook et al. (2004) study a model that can be regarded as having deteriorating jobs or impatient jobs, where an unfinished job is always subject to a transition probability to leave the system, regardless of whether it is being processed or not. Their objective is to maximize the total discounted reward (similar to other multi-armed bandit problems), so that their problem can be further modeled as a restless multiarmed bandit process (RMABP). RMABP was first introduced by Whittle (1984) and then by Niño-Mora (2001) in the context of partial conservation laws (under which a job always has finite states). It is a generalization of classical multiarmed bandit processes to allow for the transition of states of arms even when it is passive (not being engaged). Whittle's index heuristic is used in the model of Glazebrook et al. (2004), but no optimal solution for RMABP seems to be available so far. In this paper, since the objective function of our problem is the makespan instead of discounted rewards, the restless bandits model does not seem to be applicable. Hence we derive our results using a different approach.*

3 Processibility

Due to the joint impact of deterioration and breakdowns, a job may be *unprocessible* in the sense that its processing requirement is strictly positive at every time instant, so that the job will never be completed. Such a phenomenon is possible in the situations where, for example, the uptimes are too short and/or the downtimes are too long, so that the deterioration outpaces the accumulation of processing achievement. As a result, unlike in classical models, the occupying time $\mathcal{O}(x, s) = \mathcal{O}(x + \alpha s)$ may be an extended random variable that is infinite with a positive probability. We formally give the following definition.

Definition 3.1 *A job is said to be "processible" at time s for initial processing requirement x if $\Pr(\mathcal{O}(x, s) < \infty) = 1$, and the processing is said to be "regular" if $\Pr(\mathcal{O}(x) < \infty) = 1$ for all initial value x.*

It is clear that for the job to be processible, certain conditions should be satisfied by the initial processing requirement x, the deterioration rate and the breakdown process. We first deal with the simplest case where the job is processed at time zero with initial requirement x, and then extend the results to allow the job to be selected to process at an arbitrary time point. Denote $\psi(x) = \Pr(\mathcal{O}(x) < \infty)$ for a deterministic initial processing requirement x. For a random initial requirement X it is evident that the probability $\psi_X = \Pr(\mathcal{O}(X) < \infty)$ can be computed by $\psi_X = \mathrm{E}[\psi(X)]$. Therefore, we now deal with $\psi(x)$. For the time being we rearrange the up/down times as $\{(Z_1, Y_1), (Z_2, Y_2), \ldots (Z_k, Y_k), \ldots\}$. That is, the job experiences a downtime before the first uptime. This process will be referred to as a *downtime-first* process and the original one an *uptime-first* process. The downtime-first process is made from the uptime-first process by taking away the first uptime of the machine, for which the processing requirement process is shown in Figure 2 below.

Again let $\tau_k = Z_k + Y_k$. In the downtime-first case, the requirement process is

$$X(t) = x + \alpha t - \sum_{k=0}^{N(t)} Y_k - \max(t - T_{N(t)} - Z_{N(t)+1}, 0), \tag{3.1}$$

provided that the initial requirement is still x, where $T_m = \sum_{k=0}^{m} \tau_k$. In particular,

$$X(T_m) = x + \alpha T_m - \sum_{k=0}^{m} Y_k = x + \sum_{k=0}^{m} [\alpha Z_k - (1 - \alpha) Y_k]. \tag{3.2}$$

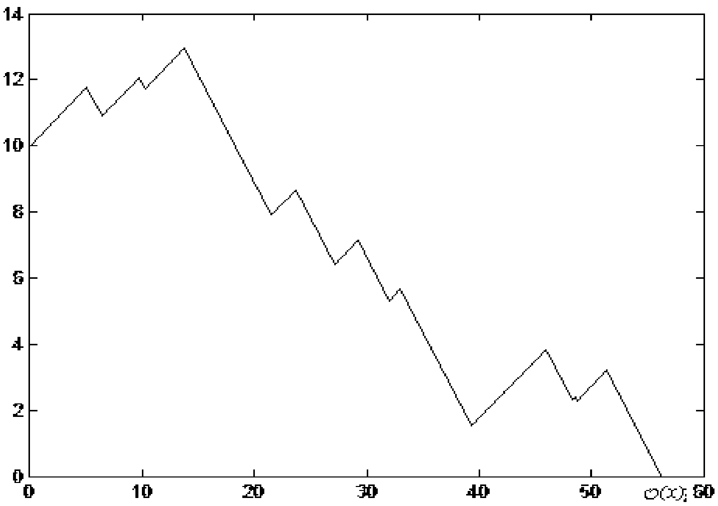

Figure 2: A typical realization of a downtime-first processing requirement

The occupying time for a downtime-first process, denoted by $\mathcal{O}_1(x)$, is also defined as such that $X(\mathcal{O}_1(x)) = 0$. Then $\mathcal{O}(x)$ can be rewritten in terms of $\mathcal{O}_1(x)$ as

$$\mathcal{O}(x) = \frac{x}{1-\alpha} \wedge Y_1 + I\left(Y_1 < \frac{x}{1-\alpha}\right) \mathcal{O}_1(x - (1-\alpha)Y_1).$$

An immediate result is

$$\mathcal{O}(x) = \infty \quad \text{if and only if} \quad Y_1 < \frac{x}{1-\alpha} \quad \text{and} \quad \mathcal{O}_1(x - (1-\alpha)Y_1) = \infty,$$

which results in the relation

$$\Pr(\mathcal{O}(x) = \infty) = \mathrm{E}\left[I\left(Y_1 < \frac{x}{1-\alpha}\right) \Pr\left(\mathcal{O}_1(x - (1-\alpha)Y_1) = \infty | \, Y_1\right)\right]. \tag{3.3}$$

So we see that the conclusion on the processibility regarding $\mathcal{O}(x)$ can be implied by the corresponding conclusions regarding $\mathcal{O}_1(x)$, and the latter is more tractable mathematically. We now turn to the calculation of the probability $\Pr\left(\mathcal{O}_1(x) < \infty\right)$ under the downtime-first process. Write $\phi(x) = \Pr(\mathcal{O}_1(x) = \infty)$. The following lemma describes the renewal equation satisfied by $\phi(x)$.

Lemma 3.1 *$\phi(x)$ satisfies the renewal equation*

$$\phi(x) = \int_x^\infty \int_0^s \phi(s-t) d\tilde{G}(t) d\tilde{H}(s), \tag{3.4}$$

where $\tilde{G}(t) = G(t/(1-\alpha))$ and $\tilde{H}(s) = H((s-x)/\alpha)$.

Proof. First notice that $\mathcal{O}_1(x) = \infty$ if and only if $X(T_n) > 0$ for all $n = 1, 2, \dots$, we see that

$$\phi(x) = \Pr(X(T_m) > 0, m = 1, 2, \dots) = \Pr(X(T_1) > 0, X(T_m) > 0,\ m = 2, 3, \dots).$$

Using the law of iterated expectations, it follows that

$$\begin{aligned}\phi(x) &= \mathrm{E}[\mathrm{E}[I(X(T_1) > 0, X(T_m) > 0,\ m = 2, 3, \dots)|X(T_1)]] \\ &= \mathrm{E}[I(X(T_1) > 0)\mathrm{E}[I(X(T_m) > 0,\ m = 2, 3, \dots)|X(T_1)]] \\ &= \mathrm{E}[I(X(T_1) > 0)\phi(X(T_1))],\end{aligned}$$

where the last equality follows from the relation

$$\mathrm{E}[I(X(T_m) > 0,\ m = 2, 3, ...)|X(T_1)] = \Pr\left(I(X(T_m) > 0,\ m = 2, 3, ...)|X(T_1)\right) = \phi(X(T_1)).$$

Using expression (3.2), we further have

$$\begin{aligned}\phi(x) &= \mathrm{E}[I(x + \alpha Z_1 - (1-\alpha)Y_1 > 0)\phi(x + \alpha Z_1 - (1-\alpha)Y_1] \\ &= \int_0^\infty \int_0^{(u+\alpha z)/(1-\alpha)} \phi(x + \alpha z - (1-\alpha)y) dG(y) dH(z). \qquad (3.5)\end{aligned}$$

Then (3.4) follows by transforming variables from (y, z) to (s, t) by $s = x + \alpha z$ and $t = (1-\alpha)y$ in the integral in (3.5). The proof is thus complete. ∎

¿From now on, we suppose that Z and Y follow exponential distributions. The next theorem gives the representation of $\phi(x)$.

Theorem 3.1 *If Z and Y follow exponential distributions with rate λ_Z and λ_Y, we have*

$$\phi(x) = \begin{cases} 1 - \dfrac{(1-\alpha)\lambda_Z}{\alpha\lambda_Y} \exp\left\{-\left(\dfrac{\lambda_Y}{1-\alpha} - \dfrac{\lambda_Z}{\alpha}\right)x\right\} & \text{if } \dfrac{\lambda_Y}{1-\alpha} > \dfrac{\lambda_Z}{\alpha}, \\ 0 & \text{otherwise.} \end{cases} \qquad (3.6)$$

Before proving Theorem 3.1 on the probability of processibility, we need a lemma on the boundary of $\phi(x)$ in the case $\lambda_Y/(1-\alpha) > \lambda_Z/\alpha$.

Lemma 3.2 *If Z and Y follow exponential distributions with rate λ_Z and λ_Y respectively, i.e., $G(u) = 1 - e^{-\lambda_Y u}$ and $H(u) = 1 - e^{-\lambda_Z u}, u \geq 0$, and $\lambda_Y/(1-\alpha) > \lambda_Z/\alpha$, then*

$$\phi(x) \geq 1 - \exp\left\{-\left(\frac{\lambda_Y}{1-\alpha} - \frac{\lambda_Z}{\alpha}\right)x\right\}.$$

Proof. We define $\phi_m(x) = \Pr(X(T_k) > 0,\ k = 1, 2, \ldots, m)$. Then it is clear that

$$\phi(x) = \lim_{m\to\infty} \phi_m(x).$$

So for the lemma, it suffices to show

$$\phi_m(x) > 1 - \exp\left\{ -\left(\frac{\lambda_Y}{1-\alpha} - \frac{\lambda_Z}{\alpha} \right) x \right\} \quad \text{for all } m \geq 1. \tag{3.7}$$

We proceed with induction arguments. For $m = 1$,

$$\phi_1(x) = \Pr(X(T_1) > 0) = \Pr(x + \alpha Z_1 - (1-\alpha)Y_1 > 0).$$

A straightforward computation shows that

$$\phi_1(x) = 1 - \frac{(1-\alpha)\lambda_Z}{(1-\alpha)\lambda_Z + \alpha\lambda_Y} \exp\left\{ -\frac{\lambda_Y}{1-\alpha} x \right\}. \tag{3.8}$$

Under the condition $\lambda_Y/(1-\alpha) > \lambda_Z/\alpha$, (3.7) follows from (3.8) for $m = 1$.

Suppose that (3.7) holds for m. We consider the case $m + 1$. Since

$$\begin{aligned} \phi_{m+1}(x) &= \Pr(X(T_k) > 0, k = 1, 2, \ldots, m+1). \\ &= \mathrm{E}\left[I(X(T_k) > 0) \Pr(X(T_k) > 0,\ k = 2, \ldots, m+1 | X(T_1)) \right] \\ &= \mathrm{E}\left[I(X(T_1) > 0) \phi_m(X(T_1)) \right]. \end{aligned}$$

By the induction hypothesis, we have

$$\phi_{m+1}(x) > \mathrm{E}\left[I(X(T_1) > 0) \left(1 - e^{-RX(T_1)} \right) \right] = \phi_1(x) - \mathrm{E}\left[I(X(T_1) > 0) e^{-RX(T_1)} \right], \tag{3.9}$$

where $R = \lambda_Y/(1-\alpha) - \lambda_Z/\alpha > 0$. Furthermore,

$$\begin{aligned} &\mathrm{E}\left[I(X(T_1) > 0) e^{-RX(T_1)} \right] = \mathrm{E}\left[I(x + \alpha Z_1 > (1-\alpha)Y_1) e^{-R(x+\alpha Z_1 - (1-\alpha)Y_1)} \right] \\ &= \lambda_Y \mathrm{E}\left[e^{-R(x+\alpha Z_1)} \int_0^{(x+\alpha Z_1)/(1-\alpha)} e^{R(1-\alpha)y - \lambda_Y y} dy \right]. \end{aligned}$$

Using $R = \lambda_Y/(1-\alpha) - \lambda_Z/\alpha$ yields

$$\mathrm{E}\left[I(X(T_1) > 0) e^{-RX(T_1)} \right] = \lambda_Y \mathrm{E}\left[e^{-R(x+\alpha Z_1)} \int_0^{(x+\alpha Z_1)/(1-\alpha)} e^{-\frac{(1-\alpha)\lambda_Z}{\alpha} y} dy \right].$$

Consequently, we see that

$$\mathrm{E}\left[I(X(T_1) > 0) e^{-RX(T_1)} \right] = e^{-Rx} - \frac{\alpha\lambda_Y}{\alpha\lambda_Y + (1-\alpha)\lambda_Z} e^{-\lambda_Y x/(1-\alpha)}.$$

Substitute this and (3.8) into (3.9), we obtain

$$\phi_{m+1}(x) > 1 - e^{-Rx} + \frac{\alpha\lambda_Y - (1-\alpha)\lambda_Z}{\alpha\lambda_Y + (1-\alpha)\,\lambda_Z} e^{-\lambda_Y x/(1-\alpha)} > 1 - e^{-Rx}.$$

So the lemma is proved by the induction principle. ∎

Proof of Theorem *3.1.* Since Z and Y follow exponential distributions with rates λ_Z and λ_Y respectively, renewal equation (3.4) becomes

$$\phi(x) = \lambda_Y\lambda_Z \int_0^\infty \int_0^{(x+\alpha z)/(1-\alpha)} \phi(x + \alpha z - (1-\alpha)y) e^{-(\lambda_Y y + \lambda_Z z)} dydz. \tag{3.10}$$

Let $s = x + \alpha z$ and $t = x + \alpha z - (1-\alpha)y$. Then the renewal equation (3.10) for $\phi(x)$ can be rewritten as

$$\phi(x) = \frac{\lambda_Y\lambda_Z}{\alpha(1-\alpha)} e^{\lambda_Z x/\alpha} \int_x^\infty \exp\left\{-\left(\frac{\lambda_Y}{1-\alpha} + \frac{\lambda_Z}{\alpha}\right) s\right\} \int_0^s \phi(t) e^{\lambda_Y t/(1-\alpha)} dtds,$$

so that

$$\phi(x)e^{-\lambda_Z x/\alpha} = \frac{\lambda_Y\lambda_Z}{\alpha(1-\alpha)} \int_x^\infty \exp\left\{-\left(\frac{\lambda_Y}{1-\alpha} + \frac{\lambda_Z}{\alpha}\right) s\right\} \int_0^s \phi(t) e^{\lambda_Y t/(1-\alpha)} dtds.$$

Differentiating the both sides of this equation with respect to x and then multiplying them with $\exp\left\{\left(\frac{\lambda_Y}{1-\alpha} + \frac{\lambda_Z}{\alpha}\right) x\right\}$, we get

$$\left[\phi'(x) - \frac{\lambda_Z}{\alpha}\phi(x)\right] e^{\lambda_Y x/(1-\alpha)} = -\frac{\lambda_Y\lambda_Z}{\alpha(1-\alpha)} \int_0^x \phi(t) e^{\lambda_Y t/(1-\alpha)} dtds. \tag{3.11}$$

By taking the second differentiation on the both sides of (3.11) with respect to x and multiplying them again by $e^{-\lambda_Y x/(1-\alpha)}$, it follows that

$$\phi''(x) + \left[\frac{\lambda_Y}{1-\alpha} - \frac{\lambda_Z}{\alpha}\right] \phi'(x) = 0. \tag{3.12}$$

This is a second-ordered differential equation. Its general solutions can be expressed as follows according to the value of $R = \lambda_Y/(1-\alpha) - \lambda_Z/\alpha$.

If $R \neq 0$, the general solution to (3.12) is, for $x > 0$,

$$\phi(x) = C_1 + C_2 e^{-Rx}. \tag{3.13}$$

As a result,

$$\phi(0+) = C_1 + C_2 \quad \text{and} \quad \phi'(0+) = C_1 - RC_2. \tag{3.14}$$

On the other hand, by (3.11),

$$\phi'(0+) = \frac{\lambda_Z}{\alpha}\phi(0+). \tag{3.15}$$

Substituting (3.14) into (3.15), we get $C_2 = -(1-\alpha)\lambda_Z C_1/\alpha\lambda_Y$. Hence

$$\phi(x) = C_1\left(1 - \frac{(1-\alpha)\lambda_Z}{\alpha\lambda_Y}e^{-Rx}\right). \tag{3.16}$$

Consider the following cases:

Case 3.1 $\frac{\lambda_Y}{1-\alpha} > \frac{\lambda_Z}{\alpha}$ *($R > 0$). Noticing that $\phi(\infty) = 1$ due to lemma 3.2, we have $C_1 = 1$. Thus*

$$\phi(x) = 1 - \frac{(1-\alpha)\lambda_Z}{\alpha\lambda_Y}\exp\left\{-\left(\frac{\lambda_Y}{1-\alpha} - \frac{\lambda_Z}{\alpha}\right)x\right\}, \quad x > 0.$$

Case 3.2 $\frac{\lambda_Y}{1-\alpha} = \frac{\lambda_Z}{\alpha}$ *($R = 0$). It is straightforward to see that $\phi(x) = 0$.*

Case 3.3 $\frac{\lambda_Y}{1-\alpha} < \frac{\lambda_Z}{\alpha}$ *($R < 0$). In this case since* $\exp\{-Rx\}$ *trends to infinity as x grows, we must have $C_1 = 0$ in order to ensure $\phi(x) \geq 0$, so that $\phi(x) = 0$ again.*

This completes the proof. ∎

Consequently, the probability $\Pr(\mathcal{O}(x) = \infty)$ can be computed by Theorem 3.1, which is presented in the next theorem.

Theorem 3.2 *If Z and Y follow exponential distributions with rate λ_Z and λ_Y, respectively, then*

$$\Pr(\mathcal{O}(x) = \infty) = \begin{cases} 1 - \exp\left\{-\left(\frac{\lambda_Y}{1-\alpha} - \frac{\lambda_Z}{\alpha}\right)x\right\} & if \quad \frac{\lambda_Y}{1-\alpha} > \frac{\lambda_Z}{\alpha} \\ 0 & if \quad \frac{\lambda_Y}{1-\alpha} \leq \frac{\lambda_Z}{\alpha} \end{cases}.$$

Proof. By (3.3) and (3.6), if $\lambda_Y/(1-\alpha) > \lambda_Z/\alpha$, then

$$\begin{aligned}
&\Pr(\mathcal{O}(x) = \infty) \\
&= \mathrm{E}\left[I\left(Y_1 < \frac{x}{1-\alpha}\right)\left[1 - \frac{(1-\alpha)\lambda_Z}{\alpha\lambda_Y}\exp\left\{-\left(\frac{\lambda_Y}{1-\alpha} - \frac{\lambda_Z}{\alpha}\right)[x - (1-\alpha)Y_1]\right\}\right]\right] \\
&= \lambda_Y\int_0^{x/(1-\alpha)}\left[1 - \frac{(1-\alpha)\lambda_Z}{\alpha\lambda_Y}\exp\left\{-\left(\frac{\lambda_Y}{1-\alpha} - \frac{\lambda_Z}{\alpha}\right)[x - (1-\alpha)y]\right\}\right]e^{-\lambda_Y y}dy \\
&= 1 - e^{-\lambda_Y x/(1-\alpha)} - \frac{(1-\alpha)\lambda_Z}{\alpha}\exp\left\{-\left(\frac{\lambda_Y}{1-\alpha} - \frac{\lambda_Z}{\alpha}\right)x\right\}\int_0^{x/(1-\alpha)} e^{-(1-\alpha)\lambda_Z y/\alpha}dy.
\end{aligned}$$

Simple computation gives

$$\Pr(\mathcal{O}(x)=\infty) = 1-\exp\left\{-\left(\frac{\lambda_Y}{1-\alpha}-\frac{\lambda_Z}{\alpha}\right)x\right\}.$$

If $\lambda_Y/(1-\alpha) \le \lambda_Z/\alpha$, then Theorem 3.1 and equation (3.3) implies $\Pr(\mathcal{O}(x)=\infty)=0$. ∎

This theorem indicates that the processibility is equivalent to the inequality

$$\frac{\lambda_Y}{1-\alpha} \le \frac{\lambda_Z}{\alpha},$$

which is independent of the initial requirement x. Therefore, for the job processing to be regular, it suffices to have a deterministic initial processing requirement x such that the job is processible. Alternatively, the condition is $\lambda_Y/(1-\alpha) \le \lambda_Z/\alpha$, or equivalently, $(1-\alpha)/\lambda_Y \ge \alpha/\lambda_Z$. That is, the capability of processing should be over the capability of deterioration regardless of the initial processing requirement.

4 The Characteristics of Occupying Time

In this section we calculate some numerical characteristics of the occupying time, including its expectation and variance, via Laplace transform. We return to the uptime-first process and consider the problem without the constraints of exponentially distributed uptimes and downtimes. Let $\varphi(x,r) = \mathrm{E}[e^{-r\mathcal{O}(x)}]$ denote the Laplace transform of the occupying time $\mathcal{O}(x)$. We first present a lemma on the renewal equation for $\varphi(x,r)$.

Lemma 4.1 *$\varphi(x,r)$ satisfies the following renewal equation:*

$$\varphi(x,r) = e^{-rx/(1-\alpha)}S_{Y_1}\left(\frac{x}{1-\alpha}-\right) + \mathrm{E}\left[e^{-r\tau_1}\varphi(x+\alpha Z_1-(1-\alpha)Y_1,r)I\left(Y_1<\frac{x}{1-\alpha}\right)\right], \tag{4.1}$$

where S_{Y_1} is the tail probability of Y_1, so that $S_{Y_1}\left(\frac{x}{1-\alpha}-\right) = \Pr\left(Y_1 \ge \frac{x}{1-\alpha}\right)$.

Proof. If the processing is finished by the first breakdown, i.e., $Y_1 \ge x/(1-\alpha)$, it is clear that $\mathcal{O}(x) = x/(1-\alpha)$. Otherwise, at the time point $\tau_1 = Y_1 + Z_1$ when the first breakdown finished and the job is to be processed again on the machine, the new processing requirement becomes $\mathcal{O}(\tau_1) = x + \alpha\tau_1 - Y_1 = x + \alpha Z_1 - (1-\alpha)Y_1$, and the new occupying time of the job from this point onwards is then $\mathcal{O}(X(\tau_1)) = \mathcal{O}(x+\alpha Z_1-(1-\alpha)Y_1)$. Combining these two cases, we can express the total occupying time $\mathcal{O}(x)$ as

$$\mathcal{O}(x) = \frac{x}{1-\alpha}I\left(Y_1 \ge \frac{x}{1-\alpha}\right) + [\tau_1 + T(X(\tau_1))]\,I\left(Y_1 < \frac{x}{1-\alpha}\right).$$

Therefore,

$$e^{-r\mathcal{O}(x)} = e^{-rx/(1-\alpha)} I\left(Y_1 \geq \frac{x}{1-\alpha}\right) + e^{-r(\tau_1 + T(x+\alpha Z_1 - (1-\alpha)Y_1))} I\left(Y_1 < \frac{x}{1-\alpha}\right), \tag{4.2}$$

and $\varphi(x, r)$ can be represented as

$$\varphi(x, r) = e^{-rx/(1-\alpha)} S_{Y_1}\left(\frac{x}{1-\alpha}-\right) + \mathrm{E}\left[e^{-r(\tau_1 + T(x+\alpha Z_1 - (1-\alpha)Y_1))} I\left(Y_1 < \frac{x}{1-\alpha}\right)\right].$$

Then (4.1) can be derived by using the law of iterated expectation and by computing the conditional expectation in the second term given Y_1. ∎

This lemma gives the renewal equation for $\varphi(x, r)$ under general conditions, which is difficult to solve in general. When Y_1 and Z_1 are exponentially distributed, however, we can get an analytic form for $\varphi(x, r)$ from the renewal equation in (4.1), which is shown in the theorem below.

Theorem 4.1 *If Y_1 and Z_1 are independent and exponentially distributed with rates λ_Y and λ_Z respectively, then*

$$\varphi(x, r) = e^{R_2(r)x}, \tag{4.3}$$

where R_2 is the non-positive root of the quadratic equation

$$R^2 + \left(\frac{\lambda_Y + r}{1-\alpha} - \frac{\lambda_Z + r}{\alpha}\right) R - \left(\frac{\lambda_Z + r}{\alpha}\frac{\lambda_Y + r}{1-\alpha} - \frac{\lambda_Y \lambda_Z}{(1-\alpha)\alpha}\right) = 0. \tag{4.4}$$

Proof. See Appendix. ∎

This theorem reveals an intuitive but important fact that the breakdowns essentially change the linear structure of the deterioration. Even though the increment of processing requirement is a linear function of the time passed, the real occupying time of the job is no longer linear with respect to the start time of the job processing, which will be clearly elaborated in what follows. This is the reason why the complexity of the scheduling problem increases dramatically when the machine is subject to breakdowns. Let $\mathcal{O}(x, s) = \mathcal{O}(x + \alpha s)$ denote the occupying time of the job when it starts at time s. Then Theorem 4.1 implies

$$\mathrm{E}[e^{-r\mathcal{O}(x,s)}] = \mathrm{E}[e^{-r\mathcal{O}(x+\alpha s)}] = e^{R_2(r)(x+\alpha s)} = e^{\alpha R_2(r)s}\mathrm{E}[e^{-r\mathcal{O}(x)}]. \tag{4.5}$$

However, if $\mathcal{O}(x, s)$ were linearly dependent on time s such that $\mathcal{O}(x, s) = \mathcal{O}(x) + \delta s$ for some δ, as in the traditional model without breakdowns, then we would have

$$\mathrm{E}[e^{-r\mathcal{O}(x,s)}] = \mathrm{E}[e^{-r\mathcal{O}(x+\alpha s)}] = e^{R_2(r)(x+\alpha s)} = e^{-\alpha sr}\mathrm{E}[e^{-r\mathcal{O}(x)}]. \tag{4.6}$$

But since $R_2(r) \neq -r$, (4.6) contradicts (4.5), and hence, the occupying time is not a linearly function of the start time s.

We now compute some features of $\mathcal{O}(x)$ in the case $\lambda_Z/\alpha > \lambda_Y/(1-\alpha)$ (which ensures the regularity of the processing, see Theorem 3.2). It is presented in the following theorem. Though the occupying time does not linearly depend on the start time as remarked above, its expectation and variance do if the initial processing requirement is a known constant.

Theorem 4.2 *Assume Y and Z are independent exponential random variables with rates λ_Y and λ_Z respectively such that $\lambda_Z/\alpha > \lambda_Y/(1-\alpha)$. Write*

$$A = \frac{\lambda_Y + \lambda_Z}{(1-\alpha)\lambda_Z - \alpha\lambda_Y} \quad \text{and} \quad B = \frac{2\lambda_Y\lambda_Z}{((1-\alpha)\lambda_Z - \alpha\lambda_Y)^3}.$$

Then $\mathrm{E}[\mathcal{O}(x)] = Ax$ *and* $\mathrm{Var}[\mathcal{O}(x)] = Bx$.

Proof. See Appendix. ∎

As a result of Theorem 4.2, it is easy to check the following corollary for stochastic initial processing requirement X.

Corollary 4.1 *Under the conditions of Theorem 4.2, for a stochastic initial processing requirement X, we have*

$$\mathrm{E}[\mathcal{O}(X)] = A\mathrm{E}[X] \quad \text{and} \quad \mathrm{Var}[\mathcal{O}(X)] = B\mathrm{E}[X] + A^2\mathrm{Var}[X],$$

where A and B are as defined in Theorem 4.2.

Proof. The first equality is straightforward and the second can be checked by the following formula: $\mathrm{Var}[\mathcal{O}(X)] = \mathrm{E}[\mathrm{Var}[\mathcal{O}(X)|X]] + \mathrm{Var}[\mathrm{E}[\mathcal{O}(X)|X]] = \mathrm{E}[BX] + \mathrm{Var}[\mathrm{A}X]$. ∎

Apparently,

$$\lim_{\lambda_Z\to\infty} A = \frac{1}{1-\alpha} \quad \text{and} \quad \lim_{\lambda_Z\to\infty} B = 0.$$

Hence,

$$\lim_{\lambda_Z\to\infty} \mathrm{E}[\mathcal{O}(X)] = \frac{\mathrm{E}[X]}{1-\alpha} \quad \text{and} \quad \lim_{\lambda_Z\to\infty} \mathrm{Var}[\mathcal{O}(X)] = \frac{\mathrm{Var}[X]}{(1-\alpha)^2}.$$

Both equations still hold if replacing $\lambda_Z \to \infty$ with $\lambda_Y \to 0$. Since $\lambda_Z \to \infty$ means that the downtimes trend to 0 and $\lambda_Y \to 0$ corresponds to infinite uptimes, in either case the model reduces to the situation without breakdowns.

Remark 4.1 *When $\lambda_Z/\alpha = \lambda_Y/(1-\alpha)$, (A.4) leads to $R_2'(0) = \infty$. Hence for any initial processing requirement $x > 0$ we have $\mathrm{E}[\mathcal{O}(x)] = -R_2'(0)x = \infty$. That is, although the job is processible in this case, its expected processing time will be infinite, which should be avoided in practice as well.*

5 Optimal Policies

In this section, we address the optimal policy (sequence) that minimizes the expected makespan of the scheduling problem. We here associate index i to the parameters such as α, λ, A, B (see Theorem 4.1), and so on, to indicate the jobs for which the parameters are referred to. For simplicity, we denote a policy by $\pi = \{1, 2, \ldots, n\}$ and by $\mathcal{O}_i(\pi), i = 1, 2, \ldots, n$ the occupying time of job i (in fact the i-th processed job) under the policy π. Further write the completion time of the i-th job as $C_i(\pi)$, or simplified to $C_i, i = 1, 2, \ldots, n$. Then it is clear that $C_{\max} = C_n(\pi)$. The following formula for the expected makespan under an arbitrary sequence policy plays an important role in deriving the optimal policy.

Theorem 5.1 *If Z_i and Y_i follow exponential distributions with rates λ_{Z_i} and λ_{Y_i} respectively, and $\lambda_{Y_i}/(1-\alpha_i) < \lambda_{Z_i}/\alpha_i$ for all $i = 1, 2, \ldots, n$, then*

$$\mathrm{E}[C_{\max}] = \sum_{k=1}^{n} \prod_{j=k+1}^{n} (\alpha_j A_j + 1) A_k \mathrm{E}[X_k]. \tag{5.1}$$

where $\prod_{j=n+1}^{n}(\alpha_j A_j + 1)$ is set to 1 by convention.

Proof. We conduct the proof by induction argument on n.

First, if $n = 1$, the clearly $C_{\max} = C_1 = \mathcal{O}_1(X)$. Thus $\mathrm{E}[C_{\max}] = \mathrm{E}[\mathcal{O}_1(X)] = A_1\mathrm{E}[X_1]$ and the conclusion in (5.1) holds.

Next, assume the induction hypothesis that (5.1) holds for $n = m$. Then we consider $n = m+1$. In this case, $C_{\max} = \mathcal{O}_{m+1}(X_{m+1} + \alpha_{m+1}C_m) + C_m$. Since, by Theorem 4.2,

$$\mathrm{E}[\mathcal{O}_{m+1}(X_{m+1} + \alpha_{m+1}C_m)] = A_{m+1}\mathrm{E}[X_{m+1} + \alpha_{m+1}C_m],$$

it follows that

$$\mathrm{E}[C_{\max}] = A_{m+1}\mathrm{E}[X_{m+1} + \alpha_{m+1}C_m] + \mathrm{E}[C_m] = A_{m+1}\mathrm{E}[X_{m+1}] + (\alpha_{m+1}A_{m+1} + 1)\mathrm{E}[C_m].$$

By the induction hypothesis, we further have

$$\begin{aligned}\mathrm{E}[C_{\max}] &= A_{m+1}\mathrm{E}[X_{m+1}] + (\alpha_{m+1}A_{m+1}+1)\sum_{k=1}^{m}\prod_{j=k+1}^{m}(\alpha_j A_j+1)A_k\mathrm{E}[X_k] \\ &= \sum_{k=1}^{m+1}\prod_{j=k}^{m+1}(\alpha_j A_j+1)A_k\mathrm{E}[X_k].\end{aligned}$$

Therefore, the theorem is proved by the induction principle. ∎

As a result, a standard interchange argument gives the following optimal policy.

Theorem 5.2 *For minimizing the expected makespan, the optimal policy orders the jobs according to nondecreasing values of* $\mathrm{E}[X_k]/\alpha_k,\ k=1,2,\ldots,n$.

Proof. We define a scheduling problem with deteriorations in the traditional sense as follows. There are n jobs which are all available at time zero and subject to deteriorations. The processing time of job k if starting at time t is $A_kX_k+\alpha_kA_kt$. Then the expected makespan for this problem is the same as in (5.1). According to the traditional results, see for example Browne and Yechiali (1990) or Alidaee and Womer (1999), the optimal policy is to sequence the jobs according to the nondecreasing order of $A_k\mathrm{E}[X_k]/\alpha_kA_k,\ k=1,2,\ldots,n$, which coincides with the nondecreasing order according to $\mathrm{E}[X_k]/\alpha_k,\ k=1,2,\ldots,n$. ∎

Remark 5.1 *Under the model formulation, as mentioned before, which includes the traditional assumption regarding linear deterioration, the optimal policy orders the jobs as if the machine has no breakdowns. This appears to be a surprising discovery, as it indicates that the breakdowns, even in a job-dependent setting, do not impact on the optimal policy to minimize expected makespan at all.*

Theorem 5.3 *Under the same conditions as in Theorem 5.1,*

$$\begin{aligned}\mathrm{Var}[C_{\max}] &= \sum_{k=1}^{n}\prod_{j=k+1}^{n}(\alpha_jA_j+1)^2\left(A_k^2\mathrm{Var}[X_k]+B_k\mathrm{E}[X_k]\right) \\ &\quad+\sum_{k=1}^{n-1}\sum_{l=k+1}^{n}\prod_{j=l+1}^{n}(\alpha_jA_j+1)^2\alpha_lB_l\prod_{j=k+1}^{l-1}(\alpha_jA_j+1)A_k\mathrm{E}[X_k]. \qquad (5.2)\end{aligned}$$

Proof. See Appendix. ∎

Note that $\lambda_{Z_i}=\infty$ (or $\lambda_{Y_i}=0$) indicates no breakdowns, the following corollary is clear from the relation $\lim_{\lambda_{Z_i}\to\infty}B_i=0$ and $\lim_{\lambda_{Z_i}\to\infty}A_i=1/(1-\alpha_i)$. It shows coincidence with the classical results on the variance of the makespan without breakdowns.

Corollary 5.1 *If no breakdowns occur, then*

$$\mathrm{Var}[C_{\max}] = \sum_{k=1}^{n} \prod_{j=k+1}^{n} \frac{1}{(1-\alpha_j)^2} \frac{\mathrm{Var}[X_k]}{(1-\alpha_k)^2}.$$

For the minimization of $\mathrm{Var}[C_{\max}]$ with machine breakdowns, however, it is difficult to construct an optimal sequence of jobs, even when all processing requirements are deterministic quantities, $x_1, x_2, \ldots, x_n$, say. In particular, we consider the simplest case with $n = 2$. Then the only sequences are $\pi_1 = \{1, 2\}$ and $\pi_2 = \{2, 1\}$. The variances of the makespan under π_1 and π_2, denoted by V_1 and V_2 respectively, are given by (5.2) as $V_1 = B_1(\alpha_2 A_2 + 1)^2 x_1 + B_2 x_2$ and $V_2 = B_2(\alpha_1 A_1 + 2)^2 x_2 + B_1 x_1$. Clearly,

$$V_1 \leq V_2 \Longleftrightarrow \frac{x_1/\alpha_1}{B_2(\alpha_1 A_1^2 + 2A_1) + B_1} \leq \frac{x_2/\alpha_2}{B_1(\alpha_2 A_2^2 + 2A_2) + B_2}.$$

In other words, even in the case $n = 2$, the optimal policy is not generally given by an index policy. This defers dramatically from the case of minimizing the expected makespan.

To understand the computational complexity more precisely, we rewrite (5.2) as

$$\begin{aligned}\mathrm{Var}[C_{\max}] =& \sum_{k=1}^{n} \prod_{j=k+1}^{n} (\alpha_j A_j + 1)^2 B_k x_k \\ &+ \sum_{k=1}^{n-1} \sum_{l=k+1}^{n} \prod_{j=l+1}^{n} (\alpha_j A_j + 1)^2 \alpha_l B_l \prod_{j=k+1}^{l-1} (\alpha_j A_j + 1) A_k x_k. \end{aligned} \tag{5.3}$$

Consider the classical scheduling problem with job deterioration defined in the proof of the optimal policy for expected makespan in Theorem 5.2. Assume that associated with each job i there is a weight w_i. Alternative to the expected makespan, we consider the total weighted expected completion time $\sum_i w_i C_i$. Browne and Yechiali (1990) showed that

$$\sum_{l=1}^{n} w_l C_l = \sum_{l=1}^{n} w_l \sum_{k=1}^{l} \prod_{j=k+1}^{l} (\alpha_j + 1) x_k = \sum_{k=1}^{n} \sum_{l=k}^{n} w_l \prod_{j=k+1}^{l} (\alpha_j + 1) x_k,$$

which appears simpler than $\mathrm{Var}[C_{\max}]$ (see (5.3)). It has been shown by Bachman et al. (2002b) that minimizing $\sum_i w_i C_i$ is an NP-hard problem. Thus we conjecture that the problem of minimizing $\mathrm{Var}[C_{\max}]$ may be NP-hard too.

6 Concluding Remarks

In this paper we have investigated the problem of scheduling a set of deteriorative jobs to minimize the expectation or the variance of the makespan subject to machine breakdowns.

We focus on the situation where the jobs experience linearly increasing deterioration and the machine is subject to preemptive-resume breakdowns. It is well known that the expectation and variance of the makespan can be minimized by index policies in the classical models without breakdowns. However, when breakdowns and deterioration are both involved, the basic features of the problems change severely. While the expected makespan still allows a simple index policy, minimizing the variance of the makespan is no longer resolvable by such a simple index policy. A more important and surprising phenomenon we have discovered is that the breakdown processes have no impact on the optimal index policy for minimizing the expected makespan.

This paper is among the first efforts to consider the simultaneous joint effects of job deterioration and machine breakdowns. Our results largely rely on the exponential distribution assumption on the up/down times. It appears to us that similar analytical results are not likely to hold under general distributions. Even for some slight departure from the exponential distribution, such as the mixed exponential distributions, the existence of similar analytical policies is not clear at all. For practical purpose, when the breakdown processes do not meet the exponential assumption, one possible approach is to use approximation: For a given distribution, we may approximate it by an exponential distribution with the same mean. The optimal index policy would be used as a heuristic rule to schedule the jobs in such a problem. In general, however, how to tackle the problem with general breakdown distributions remain an open question, which is an important topic for further studies.

Topics of future researches along this line include other patterns of machine breakdown and deterioration, such as linearly decreasing processing requirements or non-linear deterioration, and preemptive-repeat breakdowns. In particular, it would be of great interest to see if the surprising phenomenon mentioned in Remark 5.1 will remain valid when the patterns are changed or the assumptions for the model are weakened.

Acknowledgments

We wish to express our sincere gratitude to the Associate Editor and two reviewers for their constructive comments and suggestions to help improve the paper.

References

[1] Alidaee, B., Womer, N.K., 1999. Scheduling with time dependent processing times: Review and extensions. *Journal of the Operational Research Society*, **50**, 711-720.

[2] Bachman, A., Cheng, T.C.E., Janiak, A., Ng, C.T., 2002a. Scheduling start time dependent jobs to minimize the total weighted completion time. *Journal of the Operational Research Society*, **53**, 688-693.

[3] Bachman, A., Janiak, A., Kovalyov, M.Y., 2002b. Minimizing the total weighted completion time of deteriorating jobs. *Information Processing Letters*, **81**, 81-84.

[4] Biskup, D., 1999. Single-machine scheduling with learning considerations. *European Journal of Operational Research*, **115**, 173-178.

[5] Birge, J., Frenk, J. B. G., Mittenthal, J. and Rinnooy Kan, A. H. G. 1990. Single-machine scheduling subject to stochastic breakdown. *Naval Research Logistics*, **37**, 661-677.

[6] Browne, S., Yechiali, U., 1990. Scheduling deteriorating jobs on a single processor. *Operations Research*, **38**, 495-498.

[7] Cai X. and Zhou X. 1999. Stochastic scheduling on parallel machine subject to random breakdowns to minimize expected costs for earliness and tardy cost. *Operations Research*, **47**, 422-437.

[8] Cai X. and Zhou X. 2000. Asymmetric earliness-tardiness scheduling with exponential processing times on an unreliable machine. *Annals of Operations Research*, **98**, 313-331.

[9] Cai, X., Sun, X. and Zhou, X. 2003. Stochastic scheduling with preemptive-repeat machine breakdowns to minimize the expected weighted flowtime. *Probability in the Engineering and Informational Sciences*, **17**, 467-485.

[10] Cai, X., Sun, X. and Zhou, X. 2004. Stochastic scheduling subject to machine breakdowns: the preemptive-repeat model with discounted reward and other criteria. *Naval Research Logistics*, **51**, 800-817.

[11] Cai, X., Wu, X. and Zhou, X. 2005. Dynamically optimal policies for stochastic scheduling subject to preemptive-repeat machine breakdowns. *IEEE Transactions on Automation Science and Engineering*, **2**, 158-172.

[12] Cai, X., Wu, X. and Zhou, X. 2009. Stochastic Scheduling Subject to Preemptive-repeat Breakdowns with Incomplete Information. To appear in *Operations Research.*

[13] Cheng, T.C.E., Ding, Q., and Lin, B.M.T., 2004. A concise survey of scheduling with time-dependent processing times. *European Joyrnal of Operational Research*, **152**, 1-13.

[14] Frostig, E. 1991. A note on stochastic scheduling on a single machine subject to breakdown – the preemptive-repeat model. *Probability in the Engineering and Informational Sciences*, **5**, 349-354.

[15] Glazebrook, K.D. 1991. On nonpreemptive policies for stochastic single-machine scheduling with breakdowns. *Probability in the Engineering and Informational Sciences*, **5**, 77-87.

[16] Glazebrook, K.D., Assell, P.S., Dunn, R.T. and Lumley, R.R. (2004). On the optimal service to impatient tasks. *Journal of Applied Probability*, **41**, 51-72.

[17] Li, W., Braun, W.J. and Zhao, Y.Q. (1998) Stochastic Scheduling on a Repairable Machine with Erlang Uptime Distribution. *Advances in Applied Probability,* **30**, 1073-1088.

[18] Mittenthal, J. and Raghavachari, M. 1993. Stochastic single machine scheduling with quadratic early-tardy penalties. *Operations Research*, **41**, 786-796.

[19] Niño-Mora, J. (2001). Restless bandits, partial conservation laws and indexability. *Advances in Applied Probability,* **33**, 76–98.

[20] Pinedo, M. and Rammouz, E. 1988. A note on stochastic scheduling on a single machine subject to breakdown and repair. *Probability in the Engineering and Informational Sciences*, **2**, 41-49.

[21] Pinedo M. 2002. *Scheduling: Theory, Algorithms, and Systems*, 2nd Edition, Prentice Hall, Englewood Cliffs.

[22] Qi, X.D., Yin, G. and Birge, J.R. 2000a. Scheduling problems with random processing times under expected earliness/tardiness costs. *Stochastic Analysis and Applications*, **18**, 453-473.

[23] Qi, X.D., Yin, G. and Birge, J.R. 2000b. Single machine scheduling with random machine breakdowns and randomly compressible processing times. *Stochastic Analysis and Applications*, **18**, 635-653.

[24] Wei, L. and Glazebrook K.D. 1998. On stochastic machine scheduling with general distributional assumptions, *European Journal of Operational Research*, **105**, 525-536.

[25] Whittle, P.(1988). Restless bandits: Activity allocation in a changing world. *Journal of Applied Probability*, **25**, 287–298.

A Appendix

We here present the proofs for some theorems in the text.

A.1 Proof of Theorem 4.1.

Proof. Under the exponential distribution assumption for both Y and Z, by representing the expectation as an integral, renewal equation (4.1) can be rewritten as

$$\begin{aligned}\varphi(x,r) &= e^{-(\lambda_Y+r)x/(1-\alpha)} \\ &\quad + \lambda_Y\lambda_Z \int_0^\infty \int_0^\infty \varphi(x+\alpha z-(1-\alpha)y,r) I\left(y<\frac{x}{1-\alpha}\right) e^{-(r+\lambda_Y)y-(r+\lambda_Z)z} dydz.\end{aligned}$$

For this equation, let $v = x+\alpha z$ and $u = x+\alpha z-(1-\alpha)y$, and note that $y < x/(1-\alpha)$ is equivalent to $u > v-x$, we obtain

$$\begin{aligned}&\varphi(x,r) - e^{-(\lambda_Y+r)x/(1-\alpha)} \\ &= \frac{\lambda_Y\lambda_Z}{(1-\alpha)\alpha} \int_x^\infty \int_{-\infty}^v \varphi(u,r) I(u>v-x) e^{-(r+\lambda_Y)(v-u)/(1-\alpha)} e^{-(r+\lambda_Z)(v-x)/\alpha} dudv.\end{aligned}$$

Multiplying this equation by $e^{-(r+\lambda_Z)x/\alpha}$, we obtain

$$\begin{aligned}&\varphi(x,r)e^{-(r+\lambda_Z)x/\alpha} - \exp\left\{-\left(\frac{\lambda_Y+r}{1-\alpha}+\frac{\lambda_Z+r}{\alpha}\right)x\right\} \\ &= \frac{\lambda_Y\lambda_Z}{(1-\alpha)\alpha} \int_x^\infty \exp\left\{-\left(\frac{\lambda_Y+r}{1-\alpha}+\frac{\lambda_Z+r}{\alpha}\right)v\right\} \int_{v-x}^v \varphi(u,r) e^{(\lambda_Y+r)u/(1-\alpha)} dudv.\end{aligned}$$

Differentiating the equation with respect to x yields

$$\begin{aligned}&\left[\varphi'_x(x,r) - \frac{\lambda_Z+r}{\alpha}\varphi(x,r)\right] e^{-(\lambda_Z+r)x/\alpha} + \left(\frac{\lambda_Y+r}{1-\alpha}+\frac{\lambda_Z+r}{\alpha}\right)\exp\left\{-\left(\frac{\lambda_Y+r}{1-\alpha}+\frac{\lambda_Z+r}{\alpha}\right)x\right\} \\ &= \frac{\lambda_Y\lambda_Z}{(1-\alpha)\alpha}\Bigg[-\exp\left\{-\left(\frac{\lambda_Y+r}{1-\alpha}+\frac{\lambda_Z+r}{\alpha}\right)x\right\}\int_0^x \varphi(u,r)e^{(\lambda_Y+r)u/(1-\alpha)}du \\ &\qquad + \int_x^\infty \exp\left\{-\left(\frac{\lambda_Y+r}{1-\alpha}+\frac{\lambda_Z+r}{\alpha}\right)v\right\}\varphi(v-x,r)e^{(r+\lambda_Y)(v-x)/(1-\alpha)}dv\Bigg] \\ &= \frac{\lambda_Y\lambda_Z}{(1-\alpha)\alpha}\Bigg[-\exp\left\{-\left(\frac{\lambda_Y+r}{1-\alpha}+\frac{\lambda_Z+r}{\alpha}\right)x\right\}\int_0^x \varphi(u,r)e^{(r+\lambda_Y)u/(1-\alpha)}du \\ &\qquad + \exp\left\{-\left(\frac{\lambda_Y+r}{1-\alpha}+\frac{\lambda_Z+r}{\alpha}\right)x\right\}\int_0^\infty \varphi(w,r)e^{-(r+\lambda_Z)w/\alpha}dw\Bigg],\end{aligned}$$

where the last equality follows by taking $w = v - x$. Multiplying both sides of the above again by $\exp\left\{\left(\frac{\lambda_Y+r}{1-\alpha}+\frac{r+\lambda_Z}{\alpha}\right)x\right\}$ gives

$$\left[\varphi'_x(x,r)-\frac{r+\lambda_Z}{\alpha}\varphi(x,r)\right]e^{(r+\lambda_Y)x/(1-\alpha)}+\left(\frac{\lambda_Y+r}{1-\alpha}+\frac{r+\lambda_Z}{\alpha}\right)$$
$$=-\frac{\lambda_Y\lambda_Z}{(1-\alpha)\alpha}\left[\int_0^x\varphi(u,r)e^{(r+\lambda_Y)u/(1-\alpha)}du-\int_0^\infty e^{-(r+\lambda_Z)v/\alpha}\varphi(v,r)dv\right].$$

Differentiate with respect to x again and multiply by $e^{(r+\lambda_Y)x/(1-\alpha)}$, we see that

$$\varphi''_{xx}(x,r)-\frac{r+\lambda_Z}{\alpha}\varphi'_x(x,r)+\frac{r+\lambda_Y}{1-\alpha}\varphi'(x,r)-\frac{r+\lambda_Y}{1-\alpha}\frac{r+\lambda_Z}{\alpha}\varphi(x,r)=-\frac{\lambda_Y\lambda_Z}{(1-\alpha)\alpha}\varphi(x,r).$$

Rearranging the terms leads to the following equation:

$$\varphi''_{xx}(x,r)-\left(\frac{r+\lambda_Z}{\alpha}-\frac{r+\lambda_Y}{1-\alpha}\right)\varphi'_x(x,r)+\left(\frac{\lambda_Y\lambda_Z}{(1-\alpha)\alpha}-\frac{r+\lambda_Y}{1-\alpha}\frac{r+\lambda_Z}{\alpha}\right)\varphi(x,r)=0. \quad \text{(A.1)}$$

This is a second-ordered differential equation with constant coefficients and its general solution is

$$\varphi(x,r)=C_1(r)e^{R_1(r)x}+C_2(r)e^{R_2(r)x},$$

where $C_1(r)$ and $C_2(r)$ are two numbers depending on r to be determined and $R_1(r)\geq 0$ and $R_2(r)<0$ are the different roots of the corresponding characteristic equation

$$R^2+\left(\frac{\lambda_Y+r}{1-\alpha}-\frac{\lambda_Z+r}{\alpha}\right)R-\left(\frac{\lambda_Z+r}{\alpha}\frac{\lambda_Y+r}{1-\alpha}-\frac{\lambda_Y\lambda_Z}{(1-\alpha)\alpha}\right)=0. \quad \text{(A.2)}$$

Since $\mathcal{O}(x)\geq x/(1-\alpha)$, it clear that $\mathcal{O}(x)$ tends to infinity with probability 1 as $x\to\infty$. So by the dominated convergence theorem, for $r>0$,

$$\lim_{x\to\infty}\varphi(x,r)=\lim_{x\to\infty}\mathrm{E}\left[e^{-r\mathcal{O}(x)}\right]=\mathrm{E}\left[\lim_{x\to\infty}e^{-r\mathcal{O}(x)}\right]=0.$$

Therefore, $C_1(r)\equiv 0$ and so $\varphi(x,r)=C_2(r)e^{R_2(r)x}$. On the other hand, noticing that $\varphi(0,r)=1$ due to $\mathcal{O}(0)=0$, we see the $C_2(r)\equiv 1$ for all r and hence (4.3) holds. ∎

A.2 Proof of Theorem 4.2

Proof. Differentiating (4.4) with respect to r, we have

$$2R_2R_2'+\left(\frac{1}{1-\alpha}-\frac{1}{\alpha}\right)R_2+\left(\frac{\lambda_Y+r}{1-\alpha}-\frac{\lambda_Z+r}{\alpha}\right)R_2'-\frac{2r+\lambda_Y+\lambda_Z}{(1-\alpha)\alpha}=0. \quad \text{(A.3)}$$

Replacing r with 0,

$$\left(2R_2(0)+\frac{\lambda_Y}{1-\alpha}-\frac{\lambda_Z}{\alpha}\right)R_2'(0)+\left(\frac{1}{1-\alpha}-\frac{1}{\alpha}\right)R_2(0)-\frac{\lambda_Y+\lambda_Z}{(1-\alpha)\alpha}=0. \tag{A.4}$$

It is easy to check, by (4.4), that $R_2(0)=0$ when $\lambda_Z/\alpha\geq\lambda_Y/(1-\alpha)$. Hence (A.4) gives $R_2'(0)=-A$. It follows that

$$\mathrm{E}[\mathcal{O}(x)]=-\left.\frac{\partial\varphi(x,r)}{\partial r}\right|_{r=0}=-R_2'(0)x=Ax.$$

Differentiating (A.3) with respect to r once again,

$$2\left(R_2'\right)^2+2R_2R_2''+2\left(\frac{1}{1-\alpha}-\frac{1}{\alpha}\right)R_2'+\left(\frac{\lambda_Y+r}{1-\alpha}-\frac{\lambda_Z+r}{\alpha}\right)R_2''-\frac{2}{\alpha(1-\alpha)}=0. \tag{A.5}$$

Replacing r with 0,

$$2\left(\frac{\lambda_Y+\lambda_Z}{\alpha\lambda_Y-(1-\alpha)\lambda_Z}\right)^2+\frac{2}{\alpha(1-\alpha)}\frac{\alpha\lambda_Z-(1-\alpha)\lambda_Y}{\alpha\lambda_Y-(1-\alpha)\lambda_Z}+\frac{\alpha\lambda_Y-(1-\alpha)\lambda_Z}{\alpha\,(1-\alpha)}R_2''=0,$$

where the second item is the sum of the second and forth items in (A.5) with r being replaced with zero. Solving R_2'' from the above equation gives

$$R_2''(0)=\frac{2\lambda_Y\lambda_Z}{\left((1-\alpha)\lambda_Z-\alpha\lambda_Y\right)^3}=B. \tag{A.6}$$

Observe that

$$\begin{aligned}\mathrm{E}[\mathcal{O}^2(x)]&=\left.\frac{\partial^2\varphi(x,r)}{\partial r^2}\right|_{r=0}=\left.\frac{\partial^2}{\partial r^2}\left[e^{R_2(r)x}\right]\right|_{r=0}=\left[e^{R_2(r)x}\left[R_2'(r)x\right]^2+e^{R_2(r)x}R_2''(r)x\right]_{r=0}\\&=\left[R_2'(0)x\right]^2+R_2''(0)x=\mathrm{E}^2[\mathcal{O}(x)]+R_2''(0)x.\end{aligned}$$

Therefore, $\mathrm{Var}[\mathcal{O}(x)]=R_2''(0)x=Bx$. The proof is thus complete. ∎

A.3 Proof of Theorem 5.3

Proof. We prove the theorem by induction argument again.

For $n=1$, it is clear that

$$\mathrm{Var}[C_{\max}]=\mathrm{Var}[\mathcal{O}_1(X_1)]=A_1^2\mathrm{Var}[X_1]+B_1\mathrm{E}X_1.$$

which coincides with (5.2). Suppose now that (5.2) holds for $n=m$. Then consider $n=m+1$. Note that $C_{\max}=\mathcal{O}_{m+1}(X_{m+1}+\alpha_{m+1}C_m)+C_m$. Hence,

$$\begin{aligned}\mathrm{Var}[C_{\max}]=&\ \mathrm{Var}\left[\mathrm{E}\left[\mathcal{O}_{m+1}(X_{m+1}+\alpha_{m+1}C_m)+C_m\middle|\,C_m,X_{m+1}\right]\right]\\&+\mathrm{E}\left[\mathrm{Var}\left[\mathcal{O}_{m+1}(X_{m+1}+\alpha_{m+1}C_m)+C_m\middle|\,C_m,X_{m+1}\right]\right].\end{aligned}$$

Since, given C_m and X_{m+1},

$$\mathrm{E}\left[\mathcal{O}_{m+1}(X_{m+1}+\alpha_{m+1}C_m)+C_m\middle|\, C_m, X_{m+1}\right] = A_{m+1}X_{m+1} + \left(\alpha_{m+1}A_{m+1}+1\right)C_m$$

and

$$\mathrm{Var}\left[\mathcal{O}_{m+1}(X_{m+1}+\alpha_{m+1}C_m)+C_m\middle|\, C_m, X_{m+1}\right] = B_{m+1}(X_{m+1}+\alpha_{m+1}C_m),$$

we further have

$$\mathrm{Var}[C_{\max}] = \mathrm{Var}\left[A_{m+1}X_{m+1} + \left(\alpha_{m+1}A_{m+1}+1\right)C_m\right] + \mathrm{E}\left[B_{m+1}(X_{m+1}+\alpha_{m+1}C_m)\right].$$

By the independence between X_{m+1} and C_m, $\mathrm{Var}[C_{\max}]$ can be rewritten as

$$\mathrm{Var}[C_{\max}] = A_{m+1}^2\mathrm{Var}[X_{m+1}] + B_{m+1}\mathrm{E}[X_{m+1}] + \left(\alpha_{m+1}A_{m+1}+1\right)^2\mathrm{Var}\left[C_m\right] + \alpha_{m+1}B_{m+1}\mathrm{E}[C_m].$$

Substituting the induction hypothesis for $n = m$ and the formula of $\mathrm{E}[C_m]$ into the above equality, we obtain

$$\begin{aligned}
\mathrm{Var}[C_{\max}] = {} & A_{m+1}^2\mathrm{Var}[X_{m+1}] + B_{m+1}\mathrm{E}[X_{m+1}] \\
& + \left(\alpha_{m+1}A_{m+1}+1\right)^2 \sum_{k=1}^{m}\prod_{j=k+1}^{m}(\alpha_j A_j+1)^2\left(A_k^2\mathrm{Var}X_k + B_k\mathrm{E}[X_k]\right) \\
& + \left(\alpha_{m+1}A_{m+1}+1\right)^2 \sum_{k=1}^{m-1}\sum_{l=k+1}^{m}\prod_{j=l+1}^{m}(\alpha_j A_j+1)^2\alpha_l B_l\prod_{j=k+1}^{l-1}(\alpha_j A_j+1)A_k\mathrm{E}[X_k] \\
& + \alpha_{m+1}B_{m+1}\sum_{k=1}^{m}\prod_{j=k+1}^{m}(\alpha_j A_j+1)A_k\mathrm{E}[X_k].
\end{aligned}$$

Combining the first three and the last two terms respectively, we get

$$\begin{aligned}
\mathrm{Var}[C_{\max}] = {} & \sum_{k=1}^{m+1}\prod_{j=k+1}^{m+1}(\alpha_j A_j+1)^2\left(A_k^2\mathrm{Var}[X_k] + B_k\mathrm{E}[X_k]\right) \\
& + \sum_{k=1}^{m-1}\sum_{l=k+1}^{m}\prod_{j=l+1}^{m+1}(\alpha_j A_j+1)^2\alpha_l B_l\prod_{j=k+1}^{l-1}(\alpha_j A_j+1)A_k\mathrm{E}[X_k] \\
& + \alpha_{m+1}B_{m+1}\sum_{k=1}^{m}\prod_{j=k+1}^{m}(\alpha_j A_j+1)A_k\mathrm{E}[X_k] \\
= {} & \sum_{k=1}^{m+1}\prod_{j=k+1}^{m+1}(\alpha_j A_j+1)^2\left(A_k^2\mathrm{Var}[X_k] + B_k\mathrm{E}[X_k]\right) \\
& + \sum_{k=1}^{m}\sum_{l=k+1}^{m+1}\prod_{j=l+1}^{m+1}(\alpha_j A_j+1)^2\alpha_l B_l\prod_{j=k+1}^{l-1}(\alpha_j A_j+1)A_k\mathrm{E}[X_k].
\end{aligned}$$

Thus (5.2) holds for $n = m + 1$. The theorem then follows from the induction principle. ∎