Ferdowsi University of Mashhad

Stochastic Models in Probability and Statistics

Print ISSN: 3115-9370
Online ISSN: 3060-7876

SMPS

Vol. 2, No. 2, 2025

In the Name of God

Stochastic Models in Probability and Statistics (SMPS)

Volume 2, Issue 2 December 2025

ISSN-Print: 3115-9370 ISSN-Online: 3060-7876

Publisher: Faculty of Mathematical Sciences, Ferdowsi University of Mashhad
Published by: Ferdowsi University of Mashhad Press
Printing Method: Electronic

Address:
SMPS Journal, Faculty of Mathematical Sciences, Ferdowsi University of Mashhad,
P.O. Box 1159, Mashhad 91775, Iran.

Tel.: +98-51-38806222 **Fax:** +98-51-38807358
E-mail: smps@um.ac.ir
Website: https://smps.um.ac.ir

Editorial Board

This journal is published under the auspices of Ferdowsi University of Mashhad

Letter from the Editor-in-Chief

It is my great pleasure to present this issue of Stochastic Models in Probability and Statistics (SMPS). As an international, peer-reviewed, open-access journal, SMPS remains committed to advancing both the theoretical foundations and practical applications of stochastic modeling, probability, and statistics—with a particular focus on reliability and related disciplines.
SMPS publishes two issues annually and is proudly supported by the Department of Statistics, Faculty of Mathematical Sciences, Ferdowsi University of Mashhad.

The journal's mission is to promote the development and dissemination of innovative methodologies and ideas that bridge stochastic modeling in probability and statistics with real-world applications. In this spirit, SMPS welcomes high-quality contributions across a broad range of topics, including reliability theory, lifetime data analysis, stochastic modeling in industrial and systems engineering, operations research, statistical methods in information theory, and dependence modeling.

On behalf of the editorial board, I would like to express my sincere gratitude to all authors for their valuable scholarly contributions, to the reviewers for their thoughtful and constructive evaluations, and to the editorial team for their unwavering dedication and professionalism. The continued success of SMPS is made possible by the collective efforts of this dynamic community of researchers and practitioners who share a common goal: advancing the field of stochastic modeling and statistical reliability analysis.

Finally, I warmly invite readers and contributors around the world to support SMPS by submitting original research, citing published work, and sharing new ideas that help shape and expand the journal's vision.

Jafar Ahmadi
Editor-in-Chief
Stochastic Models in Probability and Statistics (SMPS)

Contents

Stochastic Models in Probability and Statistics
Vol. 2, No. 2, 2025, pp. 119–132. Research Paper

A note on the signature and dynamic signature of coherent systems

Ahmad Mirjalili[†] **and Mohammad Khanjari Sadegh**[‡*]

[†]*Department of Statistics, Maybod Branch, Islamic Azad University, Maybod, Iran*
[‡]*Department of Statistics, Faculty of Mathematical Sciences, Ferdowsi University of Mashhad, P. O. Box 1159, Mashhad 91775, Iran*

Email(s): a.mirjalili@maybodiau.ac.ir, mkhanjari@um.ac.ir

Abstract. This paper investigates new properties and applications of the system signature and its dynamic counterpart, which serve as effective tools for analyzing stochastic ordering and aging properties in both coherent and used systems. For coherent systems with exchangeable components, the system signature offers a more powerful comparative framework than the traditional structure function. In the context of used systems with different exchangeable components, we propose an alternative to the dynamic signature that is simpler to implement and often preferable to the standard dynamic signature. This alternative also proves useful in scenarios where the traditional dynamic signature is inapplicable. Additionally, by examining all 28 coherent systems of order $n \leq 4$, we establish a unique property of series systems: under both identical and non-identical independent component lifetimes, series systems are the only ones that are decreasing failure rate (DFR) closed. The results extend several existing findings related to system signatures and their dynamic versions. Illustrative examples are provided to demonstrate the practical relevance of the theoretical results.

Keywords: Coherent systems, Dynamic signature, DFR, IFR, Signature, Stochastic ordering, Used systems.

1 Introduction

One of the most important problems in system reliability analysis is comparison among systems. The most common tools to compare the coherent systems are structure and reliability functions, system signatures, and distortion functions. Assume that the systems are coherent (see, Barlow and Proschan (1975) for details on coherent systems) and suppose that the component lifetimes of the system are independent and identically distributed (IID) or they are exchangeable (EXC) random variables, defined in the next section. This paper is concerned to some more properties and applications of the signatures in system comparisons. Let $T_1, \ldots, T_n$ and $T = \phi(T_1, \ldots, T_n)$ be the component lifetimes and the system lifetime, respectively. ϕ refers to the structure of the system. When T_i's are continuous and IID random variables,

*Corresponding author
Received: 5 March 2025 / Accepted: 29 September 2025
DOI: 10.22067/smps.2025.92495.1043

the following well-known and important result is obtained by Samaniego (1985)

$$P(T>t)=\sum_{i=1}^{n} s_i P(T_{i:n}>t), \tag{1}$$

where $T_{i:n}$ is the i-th ordered components lifetime, and $s_i=P(T=T_{i:n})$. The probability vector $\mathbf{s}=(s_1,\ldots,s_n)$ is called the system signature.

Using the system signature, many researchers have studied various aspects of system reliability. In fact, the identity (1) was first proved by Kochar et al. (1999). Navarro et al. (2005) claimed that this equation holds true if T_i's have an absolutely continuous exchangeable joint distribution. Samaniego (2007) published a book on the system signatures and their applications. Also Naqvi et al. (2022) have done a review and bibliometric analysis on system signatures. Navarro et al. (2008) showed that (1) holds true if T_i's be continuous (or discrete) weakly exchangeable random variables and also $P(T_i=T_j)=0$ for all $i\neq j$ (that is T_i's have no tie). It seems that the strongest result is given by Marichal et al. (2011). They proved that (1) holds true if and only if for $i=1,\ldots,n$, the following binary random variables are exchangeable

$$X_i(t)=\begin{cases} 1 & T_i>t \\ 0 & T_i\leq t. \end{cases} \tag{2}$$

Navarro et al. (2011) investigated the signatures of coherent systems with heterogeneous and independent components. Rao and Naqvi (2023) studied the stochastic comparisons of coherent systems with heterogeneous and dependent components having proportional reversed failure rates. In the EXC case, a necessary and sufficient condition for comparing two systems was given by Navarro and Rubio (2011). For more details and an extensive study on comparisons of coherent systems see Navarro (2018) and references therein.

Remark 1. *The system signature has the following interesting and important property. It should be noted that in all cases where the Equation (1) holds true (except when the component lifetimes are continuous and IID random variables), the signature is not necessarily defined by $s_i=P(T=T_{i:n})$. But in these cases, the value of s_i used in (1), is exactly determined by $s_i=P(T=T_{i:n})$ when T_i's are continuous and IID random variables. Therefore in all cases, the system signature $\mathbf{s}=(s_1,\ldots,s_n)$ remains as a probability vector. For further clarification, consider $T=min\{T_1,max(T_2,T_3)\}$, where T_i's are continuous and IID. We have $\mathbf{s}=(1/3,2/3,0)$ and $P(T=T_{i:n})=s_i$. Now suppose T_i's are discrete and IID with common distribution Binomial(1,1/2). In this case we have*

$$P(T=T_{1:3})=1-P(T_1=1,T_2=0,T_3=1)-P(T_1=1,T_2=1,T_3=0)=6/8\neq s_1=1/3.$$

But one can easily verify that for

$$P(T=x)=\sum_{1}^{3} s_i P(T_{i:n}=x),\ x=0,1.$$

That is the mixture representation (1) holds true.

The concept of the dynamic signature of the system was first defined by Samaniego et al. (2009). It is a common tool in comparisons among coherent used systems. They showed in the case of IID that

$$P(T=T_{k:n}|T>t,N(t)=r)=\frac{s_k}{\sum_{j=r+1}^{n} s_j}=w_k,$$

for $k=r+1,\ldots,n$ and $r=0,1,\ldots,n-1$. The probability vector $\mathbf{w}_{(n-r)}=(w_{r+1},\ldots,w_n)$ is called the dynamic signature of the system at time t. $N(t)$ is the number of failed components of the system up to time t and such a system is said to be a used system. Based on the dynamic signature, they obtained some ordering results for comparing coherent used systems with IID components. Note that if $r = 0$, the dynamic signature reduces to the usual signature as $\sum_1^n s_i = 1$.

Burkschat and Navarro (2013) considered dynamic signatures of coherent systems based on sequential order statistics. Under some partial information about the failure status of the system lifetime, Mahmoudi and Asadi (2011) defined the dynamic signature of coherent systems. Burkschat and Samaniego (2018) examined the concepts of dynamic and conditional increasing failure rates properties in coherent systems. Chahkandi et al. (2015) studied the signature of the repairable systems.

The remainder of this paper is organized as follows. For the sake of completeness, Section 2 provides an extensive review of comparisons among coherent systems based on structure functions, reliability functions, and system signatures. As a minor result, it is shown that, in the case of exchangeable components (EXC), the system signature is a more powerful tool than the structure function. Section 3 focuses on comparisons of coherent used systems. Using the dynamic signature, new stochastic ordering results are derived for comparing coherent used systems with different exchangeable components. A new, simple alternative to the dynamic signature is proposed, which is particularly useful in cases where the standard dynamic signature is not applicable. Due to its ease of use, this alternative should be considered as a first option in such comparisons. Finally, Section 4 examines properties related to the preservation of reliability aging classes in series systems.

Motivated by the result of Samaniego (1985) and by verifying the signatures of all 28 coherent systems of order $n \leq 4$, it is shown that when the lifetimes of the system components are either IID or independent and not identical (INID), the series systems are only decreasing the failure rate (DFR) closed systems. The increasing failure rate (IFR) closedness property of the series systems is also considered. Recall that a random variable X with absolutely continuous distribution function F and density function f has an IFR (DFR) distribution if its failure rate function $h(t) = f(t)/\bar{F}(t)$ is increasing (decreasing) in t, $\bar{F}(t) = 1 - F(t)$. Throughout the paper, when two systems are said to be ordered, it means that their lifetimes are ordered stochastically. Recall that X is less than Y in the usual stochastic order and denote it by $X \leq_{st} Y$ if $\bar{F}_X(t) \leq \bar{F}_Y(t)$ for all $t > 0$.

2 Comparisons of coherent systems

The comparison of two systems can be done at a fixed point in time (static comparison) or during periods of time (dynamic comparison). In order to present the main results of the paper given in the two next sections, and for the sake of completeness, this section mainly deals with a review of the basic results in the literature on comparisons among coherent systems. Some minor contributions are also included. At a fixed point in time, and based on the system structure and reliability functions, we first consider the comparison of two systems. In comparison among the systems with INID components, it is shown that the structure and reliability functions are two equivalent tools. Finally the system signature is used to compare two systems during of times. In the sequel, Lemma 1 shows in IID and EXC cases that the system signature is a more powerful tool than the structure-function.

At a fixed point in time, assume that binary random states are as follows:

$$X_i = \begin{cases} 1 & \text{if } i\text{th component is working} \\ 0 & \text{otherwise,} \end{cases}$$

and

$$\phi(\mathbf{X}) = \phi(X_1, \ldots, X_n) = \begin{cases} 1 & \text{if the system is working} \\ 0 & \text{otherwise.} \end{cases}$$

Their reliabilities are also defined as:

$$p_i = E(X_i) = P(X_i = 1), \; i = 1, 2, \ldots, n$$

$$h(\mathbf{p}) = h(p_1, \ldots, p_n) = E(\phi(\mathbf{X})) = P(\phi(\mathbf{X}) = 1).$$

$\phi(\mathbf{x})$ is called the structure function of the system. In addition, the order of the system is defined as the number of the system components n.

Definition 1. *(Kochar et al., 1999) Let $\phi_1(\mathbf{x})$ and $\phi_2(\mathbf{x})$ be the structure functions of two systems of order n. The second system is said to be better than the first one if*

$$\phi_1(\mathbf{x}) \leq \phi_2(\mathbf{x}) \quad \forall\, \mathbf{x} \in \{0,1\}^n. \tag{3}$$

If $\phi_1(\mathbf{x}) = \phi_2(\mathbf{x})$, $\forall \mathbf{x} \in \{0,1\}^n$ two systems are obviously the same. Therefore (3) should be strictly hold at least for one $\mathbf{x}$. The second system is said to be more reliable than the first one if

$$h_1(\mathbf{p}) = E(\phi_1(\mathbf{X})) \leq h_2(\mathbf{p}) = E(\phi_2(\mathbf{X})) \quad \forall\, \mathbf{p} \in [0,1]^n. \tag{4}$$

Again the inequality should be strict at least for one $\mathbf{p}$, otherwise two systems have the same reliability.

It is claimed in Kochar et al. (1999) that the inequalities (3) and (4) are equivalent when X_i's are independent (INID case). Note that in this case

$$h(\mathbf{p}) = P(\phi(\mathbf{X}) = 1) = \sum_{\mathbf{x}:\phi(\mathbf{x})=1} \prod_{i=1}^{n} p_i^{x_i}(1-p_i)^{1-x_i} = \sum_{\mathbf{x}:\phi(\mathbf{x})=1} \prod_{i\in\mathbf{1_x}} p_i \prod_{i\in\mathbf{0_x}} (1-p_i),$$

where $\mathbf{1_x} = \{1 \leq i \leq n | x_i = 1\}$, $\mathbf{0_x} = \{1 \leq i \leq n | x_i = 0\}$. An argument for their claim is as follows. The inequality (3) simply implies (4). Now suppose (4) holds true. If $\phi_1(\mathbf{x}) = \phi_2(\mathbf{x})$ for all $\mathbf{x}$ obviously $h_1(\mathbf{p}) = h_2(\mathbf{p})$ for all $\mathbf{p}$ which is a contradiction. Now suppose there exists a vector $\mathbf{x}_0$ such that $\phi_1(\mathbf{x}_0) = 1 > \phi_2(\mathbf{x}_0) = 0$. Put $\mathbf{p}_0 = \mathbf{x}_0$ then $p_{0i} = 1$ for $i \in \mathbf{1}_{\mathbf{x}0}$ and $1 - p_{0i} = 1$ for $i \in \mathbf{0}_{\mathbf{x}0}$. Note that $P_{\mathbf{p}_0}(\mathbf{X} = \mathbf{x}) = P(\mathbf{X} = \mathbf{x}|\mathbf{p} = \mathbf{p}_0) = 0$ for $\mathbf{x} \neq \mathbf{x}_0$. It is easy to see that $h_1(\mathbf{p}_0) = 1$ and $h_2(\mathbf{p}_0) = 0$ which is again a contradiction and this shows that (4) implies (3).

The inequalities (3) and (4) are not necessarily equivalent if X_i's are IID that is $p_i = p$, $i = 1, \ldots, n$. See the following example.

Example 1. (Kochar et al., 1999). Suppose $\phi_1(\mathbf{x}) = \min\{x_1, \max(x_2, x_3)\}$ and $\phi_2(\mathbf{x}) = \min\{x_2, \max(x_1, x_3)\}$ then $\phi_1(1,0,1) = 1 > \phi_2(1,0,1) = 0$ and $\phi_1(0,1,1) = 0 < \phi_2(0,1,1) = 1$ whereas $h_1(p) = 2p^2 - p^3 = h_2(p)$, for all $0 \leq p \leq 1$. That is (4) does not imply (3).

Remark 2. *Note that in INID case, the inequalities (3) and (4) are equivalent even if $\mathbf{p} \in (0,1)^n$. It is because as the system components are independent, the reliability function $h(\mathbf{p})$ is a multilinear, that is, linear in each p_i and therefore is a continuous function.*

Now by using the system signature, the comparison of systems during of times is considered. The first result is given by Kochar et al. (1999) as follows:
Let $T_1 = \phi_1(X_1, \ldots, X_n)$ and $T_2 = \phi_2(X_1, \ldots, X_n)$ denote the lifetimes of two coherent systems where X_i is the lifetime of the ith component and X_i's are IID and continuous and suppose $\mathbf{s}_1$ and $\mathbf{s}_2$ are their signatures respectively. If $\mathbf{s}_1 \leq_{st} \mathbf{s}_2$, then $T_1 \leq_{st} T_2$. If $\mathbf{s}_1 \leq_{hr} \mathbf{s}_2$, then $T_1 \leq_{hr} T_2$ and if $\mathbf{s}_1 \leq_{lr} \mathbf{s}_2$, then $T_1 \leq_{lr} T_2$.
The converse of their results is not true, see Rychlik et al. (2018). (We also refer to Shaked and Shanthikumar (2007) for details on stochastic orders).

Example 2. Consider two coherent systems mentioned in Example 1 where $\mathbf{s}_1 = (1/3, 2/3, 0) = \mathbf{s}_2$ and hence from (1), T_1 and T_2 are identically distributed. Note that these two systems could not be ordered by using of their structure functions.

The comparison of coherent systems with EXC components is studied by Navarro et al. (2005). Recall that the random variables $T_1,\dots,T_n$ are said to be exchangeable if $P(T_1>t_1,\dots,T_n>t_n)=P(T_{\pi(1)}>t_1,\dots,T_{\pi(n)}>t_n)$ for any permutation $\pi=(\pi(1),\dots,\pi(n))$ of numbers $\{1,\dots,n\}$.

The relationship between the signature and structure function of a system is obtained by Boland (2001) or Marichal et al. (2011) as follows. In EXC case and for $k=1,\dots,n$

$$\sum_{i=n-k+1}^{n} s_i = P(\phi(\mathbf{X})=1|\sum X_i=k)=\sum_{\mathbf{x};\sum x_i=k}\phi(\mathbf{x})\frac{1}{\binom{n}{k}}. \tag{5}$$

Lemma 1. *If $\phi_1(\mathbf{x})\le\phi_2(\mathbf{x})$ for all $\mathbf{x}\in\{0,1\}^n$ then $\mathbf{s}_1\le_{st}\mathbf{s}_2$.*

Proof. From Equation (5) we see that if $\phi_1(\mathbf{x})\le\phi_2(\mathbf{x})$ for all $\mathbf{x}\in\{0,1\}^n$ then $\sum_{i=n-k+1}^{n}s_{1,i}\le\sum_{i=n-k+1}^{n}s_{2,i}$ for $k=1,2,\dots,n$, that is $\mathbf{s}_1\le_{st}\mathbf{s}_2$. □

In fact Lemma 1 shows that the system signature is a more stronger tool than the structure function. This section is ended by reviewing the problem of comparison of coherent systems with different sizes.

Definition 2. *(Boland and Samaniego, 2004), or (Samaniego, 2006). Let $\mathbf{r}=(r_1,\dots,r_k)$ be a probability vector and suppose $T_j=\phi_j(X_1,\dots,X_n)$, $j=1,\dots,k$ are the lifetimes of k coherent systems with IID or EXC components. If $P(T=T_j)=r_j$ then T is said to be the lifetime of a mixed-$\mathbf{r}$ system. Note that in general, a mixed system is not coherent, and the class of coherent systems is a subset of the mixed systems. Let $\mathbf{s}_j$ be the signature of the jth coherent system, then $\mathbf{s}=\sum_{j=1}^{k}r_j\mathbf{s}_j$ is defined as the signature of the mixed-$\mathbf{r}$ system.*

For a system of order n, an equivalent system of order $n+1$ is introduced by Samaniego (2006) as follows.

Let T be the lifetime of a coherent or a mixed system of order n with IID components having common distribution F and with signature $\mathbf{s}=(s_1,\dots,s_n)$. Then $T=^{st}T^*$ where T^* is the lifetime of a mixed system of order $n+1$ with IID components which have the same distribution F and its signature is $\mathbf{s}^*=(s_1^*,\dots,s_{n+1}^*)$ with

$$s_i^*=\frac{(i-1)s_{i-1}+(n-i+1)s_i}{n+1}, i=1,\dots,n+1.$$

Assume that $s_0=s_{n+1}=0$. $\mathbf{s}^*$ is of order $n+1$ and is the equivalent of $\mathbf{s}$.

For a system of order k, an equivalent system of order $n(>k)$ can be defined. Note that the above result also holds true for EXC components, also note that the equivalent systems belong in general to the class of mixed systems and all comparison results among the coherent systems also hold true for the mixed systems.

An extended result for coherent systems consisting of EXC components, is given by Navarro et al. (2008) as follows.

Let $T_1=\phi_1(Y_1,\dots,Y_{n_1})$ and $T_2=\phi_2(Z_1,\dots,Z_{n_2})$ be the lifetimes of two systems where Y_i's and Z_i's are subsets of the EXC random variables $\{X_1,\dots,X_n\}$ and suppose $\mathbf{s}_1(n)$ and $\mathbf{s}_2(n)$ are their equivalent signatures of order n, respectively.

(a) If $\mathbf{s}_1(n)\le_{st}\mathbf{s}_2(n)$, then $T_1\le_{st}T_2$.

(b) If $\mathbf{s}_1(n)\le_{hr}\mathbf{s}_2(n)$ and $X_{i:n}\le_{hr}X_{i+1:n}$, $i=1,\dots,n-1$, then $T_1\le_{hr}T_2$.

(c) If $\mathbf{s}_1(n)\le_{lr}\mathbf{s}_2(n)$ and $X_{i:n}\le_{lr}X_{i+1:n}$, $i=1,\dots,n-1$, then $T_1\le_{lr}T_2$.

For an extensive study on the existence and comparisons of equivalent systems with IID components refer to Lindqvist et al. (2016). It is pointed out there, within the class of mixed systems, one can always find equivalent mixed systems of larger sizes, but not necessary if the size is decreased.

3 Comparisons of used systems

This section begins by considering the comparison of coherent used systems, extends some existing results, and then proposes an alternative for equivalent dynamic signature. Its properties and applications are illustrated. The following result is given in Samaniego et al. (2009).

Let $\mathbf{s}_1$ and $\mathbf{s}_2$ be the signatures of two coherent systems of order n and with IID components having common continuous distribution F and let $T_1 = \phi_1(X_1,\ldots,X_n)$ and $T_2 = \phi_2(X_1,\ldots,X_n)$ be their lifetimes, respectively. Suppose $T_1 > t$, $T_2 > t$, $N_1(t) = r_1$ and $N_2(t) = r_2$. Let $\mathbf{w}_1 = (w_{1,1},\ldots,w_{1,n-r_1})$ and $\mathbf{w}_2 = (w_{2,1},\ldots,w_{2,n-r_2})$ be their dynamic signatures. Also assume that $\mathbf{w}_1^{(n)}$ and $\mathbf{w}_2^{(n)}$ are their equivalent versions of order n. If $\mathbf{w}_1^{(n)} \leq_{st} \mathbf{w}_2^{(n)}$ then

$$(T_1 - t|T_1 > t, N_1(t) = r_1) \leq_{st} (T_2 - t|T_2 > t, N_2(t) = r_2).$$

As defined in Section 1, $N(t)$ is the number of failed components of the system up to time t.
Samaniego et al. (2009) also provide the following expression:

$$P(T-t > x|T > t, N(t) = r) = \sum_{i=1}^{n-r} w_{r+i}\bar{F}_{i:n-r|t}(x) = \sum_{i=1}^{n} w_i^{(n)}\bar{F}_{i:n|t}(x),$$

where $\bar{F}_{i:n-r|t}(x)$ is the reliability function of the ith order statistic in a random sample of size $n-r$ from $\bar{F}_t(x) = \frac{\bar{F}(x+t)}{\bar{F}(t)}$ for $x > 0$.
Note that $N(t)$ is distributed as Binomial$(n, F(t))$ and $N(t) = r$ is equivalent to $X_{r:n} < t < X_{r+1:n}$. Therefore $P(X_{i:n} - t > x|N(t) = r) = 0$, for $i = 1,\ldots,r$ and for $i = r+1,\ldots,n$, we have

$$\begin{aligned}
P(X_{i:n} - t > x|N(t) = r) &= \frac{\binom{n}{r}F^r(t)\sum_{j=0}^{i-r-1}\binom{n-r}{j}[\bar{F}(t) - \bar{F}(t+x)]^j(\bar{F}(t+x))^{n-r-j}}{\binom{n}{r}F^r(t)(\bar{F}(t))^{n-r}} \\
&= \sum_{j=n-i+1}^{n-r}\binom{n-r}{j}(\bar{F}_t(x))^j(1-\bar{F}_t(x))^{n-r-j} \\
&= P(\text{Binomial}(n-r, \bar{F}_t(x)) \geq n-i+1). \qquad (6)
\end{aligned}$$

Now suppose $T = \phi(X_1,\ldots,X_n)$ is the lifetime of a system with EXC components. For this system, the following equation is given by Sadegh (2016)

$$\begin{aligned}
P(T-t > x|T > t, N(t) = r) &= \sum_{k=r+1}^{n} w_k P(X_{k:n} - t > x|N(t) = r) \\
&= \sum_{k=1}^{n} w_k^{(n)} P(X_{k:n} - t > x|N(t) = 0). \qquad (7)
\end{aligned}$$

Based on the Equation (7), the next theorem extends part (a) of Theorem 2.6 in Samaniego et al. (2009), to the systems with different components.

Theorem 1. *Let $T_1 = \phi_1(X_1,\ldots,X_n)$ and $T_2 = \phi_2(Y_1,\ldots,Y_n)$ be the lifetimes of two coherent systems with EXC components. Suppose $T_1 > t$, $T_2 > t$, $N_1(t) = r_1$ and $N_2(t) = r_2$. Let $\mathbf{w}_1$, $\mathbf{w}_2$, $\mathbf{w}_1^{(n)}$ and $\mathbf{w}_2^{(n)}$ are defined as before. If $\mathbf{w}_1^{(n)} \leq_{st} \mathbf{w}_2^{(n)}$ and $(X_{i:n} - t|N_1(t) = 0) \leq_{st} (Y_{i:n} - t|N_2(t) = 0)$ for $i = 1,\ldots,n$ then*

$$(T_1 - t|T_1 > t, N_1(t) = r_1) \leq_{st} (T_2 - t|T_2 > t, N_2(t) = r_2).$$

Proof. It is known that if $X \leq_{st} Y$ then $Eg(X) \leq Eg(Y)$ for any increasing function g. Let $\mathbf{w}_1^{(n)}$ and $\mathbf{w}_2^{(n)}$ be the probability vectors of two discrete random variables X and Y, respectively. Note that $X \leq_{st} Y$ as $\mathbf{w}_1^{(n)} \leq_{st} \mathbf{w}_2^{(n)}$. Now in view of the second equality of Equation (7) we have

$$\begin{aligned}
E[g(X)] &= \sum_{1}^{n} w_{1,k}^{(n)} P(X_{k:n} - t > x | N_1(t) = 0) = P(T_1 - t > x | T_1 > t, N_1(t) = r_1) \\
&\leq Eg(Y) = \sum_{k=1}^{n} w_{2,k}^{(n)} P(X_{k:n} - t > x | N_1(t) = 0) \\
&\leq \sum_{k=1}^{n} w_{2,k}^{(n)} P(Y_{k:n} - t > x | N_2(t) = 0) \\
&= P(T_2 - t > x | T_2 > t, N_2(t) = r_2).
\end{aligned}$$

□

Here is an example that satisfies the conditions of the Theorem 1.

Example 3. Let $T_1 = \phi_1(X_1, X_2, X_3) = \min\{X_1, \max(X_2, X_3)\}$ and suppose $T_2 = \phi_2(Y_1, Y_2, Y_3) = \max Y_i$. We consider two following cases.

(i) IID case: Suppose X_i's and Y_i's are independent and distributed as $\bar{F}(x) = e^{-\lambda x}$ and $\bar{G}(x) = e^{-\theta x}$, respectively and $\lambda > \theta$. Assume that $N_1(t) = N_2(t) = 1$, $T_1 > t$ and $T_2 > t$. Now these two systems are satisfying the conditions of Theorem 1 as we have:
$\mathbf{s}_1 = (1/3, 2/3, 0)$, $\mathbf{s}_2 = (0, 0, 1)$ and therefore $\mathbf{w}_1 = (1, 0)$ and $\mathbf{w}_2 = (0, 1)$.
Also $\mathbf{w}_1^{(3)} = (2/3, 1/3, 0)$ and $\mathbf{w}_2^{(3)} = (0, 1/3, 2/3)$. Obviously $\mathbf{w}_1^{(3)} \leq_{st} \mathbf{w}_2^{(3)}$ and

$$\bar{F}_t(x) = \bar{F}(x) = e^{-\lambda x} < \bar{G}_t(x) = \bar{G}(x) = e^{-\theta x}.$$

It is known that the Binomial distribution $B(n, p)$ is stochastically increasing in p. Therefore from Equation (6), we have $P(X_{i:n} - t > x | N_1(t) = 0) \leq P(Y_{i:n} - t > x | N_2(t) = 0)$ for $i = 1, 2, 3$. It implies that

$$(T_1 - t | T_1 > t, N_1(t) = 1) \leq_{st} (T_2 - t | T_2 > t, N_2(t) = 1).$$

For more clarity, we compare their survival functions as follows. From structures of these two systems we have

$$\begin{aligned}
P(T_1 - t > x | T_1 > t, N_1(t) = 1) &= \frac{2P(X_1 > t + x, X_2 > t + x, X_3 < t)}{2P(X_1 > t, X_2 > t, X_3 < t)} \\
&= \frac{[P(X_1 > t + x)]^2}{[P(X_1 > t)]^2} \\
&= \frac{e^{-2\lambda(t+x)}}{e^{-2\lambda t}} = e^{-2\lambda x}.
\end{aligned}$$

Now for the second system we have

$$\begin{aligned}
P(T_2 - t > x | T_2 > t, N_2(t) = 1) &= \frac{6P(X_1 > t + x, t < X_2 < t + x, X_3 < t) + 3P(X_1 > t + x, X_2 > t + x, X_3 < t)}{3P(X_1 > t, X_2 > t, X_3 < t)} \\
&= \frac{6e^{-\theta(t+x)}(e^{-\theta t} - e^{-\theta(t+x)}) + 3e^{-2\theta(t+x)}}{3e^{-2\theta t}} \\
&= 2e^{-\theta x} - e^{-2\theta x}.
\end{aligned}$$

It is easy to see that

$$e^{-2\lambda x} \leq 2e^{-\theta x} - e^{-2\theta x},$$

as $\lambda > \theta$.

(ii) EXC case: Let $\bar{F}(x_1,x_2,x_3) = e^{-\Sigma_1^3 x_i - \lambda x_{3:3}}$ and $\bar{G}(y_1,y_2,y_3) = e^{-\Sigma_1^3 y_i - \theta y_{3:3}}$ where $\lambda > \theta$ and $x_{3:3} = \max(x_1,x_2,x_3)$. Again assume that $N_1(t) = N_2(t) = 1$, $T_1 > t$ and $T_2 > t$. We now verify the conditions of Theorem 1.
If $i = 1$, then

$$\begin{aligned}
P(X_{1:3} - t > x | N_1(t) = 0) &= P(X_{1:3} > t + x)/P(X_{1:3} > t) \\
&= \bar{F}(t+x,t+x,t+x)/\bar{F}(t,t,t) \\
&= e^{-3t-3x-\lambda(t+x)}/e^{-3t-\lambda t} = e^{-3x-\lambda x} \\
&\leq e^{-3x-\theta x} \\
&= P(Y_{1:3} - t > x | N_2(t) = 0).
\end{aligned}$$

That is $(X_{1:3} - t | N_1(t) = 0) \leq_{st} (Y_{1:3} - t | N_2(t) = 0)$. If $i = 2$, then from exchangeability of X_i's we have

$$\begin{aligned}
P(X_{2:3} - t > x | N_1(t) = 0) &= \frac{\bar{F}(t+x,t+x,t+x) + 3P(t < X_1 < t+x, X_2 > t+x, X_3 > t+x)}{\bar{F}(t,t,t)} \\
&= \frac{e^{-3t-3x-\lambda(t+x)} + 3(\bar{F}(t,t+x,t+x) - \bar{F}(t+x,t+x,t+x))}{e^{-3t-\lambda t}} \\
&= \frac{3e^{-3t-2x-\lambda(t+x)} - 2e^{-3t-3x-\lambda(t+x)}}{e^{-3t-\lambda t}} = e^{-\lambda x-2x}(3 - 2e^{-x}) \\
&\leq e^{-\theta x-2x}(3 - 2e^{-x}) \\
&= P(Y_{2:3} - t > x | N_2(t) = 0).
\end{aligned}$$

That is $(X_{2:3} - t | N_1(t) = 0) \leq_{st} (Y_{2:3} - t | N_2(t) = 0)$.
Similarly it can be shown that $(X_{3:3} - t | N_1(t) = 0) \leq_{st} (Y_{3:3} - t | N_2(t) = 0)$. Hence all conditions of the Theorem 1 are satisfied and therefore

$$(T_1 - t | T_1 > t, N_1(t) = 1) \leq_{st} (T_2 - t | T_2 > t, N_2(t) = 1).$$

For more clarity, we also obtain the corresponding survival functions to see that the above stochastic ordering between these two systems in fact holds true. For the first system we have

$$\begin{aligned}
P(T_1 - t > x | T_1 > t, N_1(t) = 1) &= \frac{2P(X_1 > t+x, X_2 > t+x, X_3 < t)}{2P(X_1 > t, X_2 > t, X_3 < t)} \\
&= \frac{\bar{F}(t+x,t+x,0) - \bar{F}(t+x,t+x,t)}{\bar{F}(t,t,0) - \bar{F}(t,t,t)} \\
&= \frac{e^{-2(t+x)-\lambda(t+x)} - e^{-3t-2x-\lambda(t+x)}}{e^{-2t-\lambda t} - e^{-3t-\lambda t}} = e^{-\lambda x-2x}.
\end{aligned}$$

Now for the second system we have

$$
\begin{aligned}
P(T_2-t>x|T_2>t,N_2(t)=1) &= \frac{6P(X_1>t+x,t<X_2<t+x,X_3<t)+3P(X_1>t+x,X_2>t+x,X_3<t)}{3P(X_1>t,X_2>t,X_3<t)} \\
&= \frac{6[\bar{F}(t+x,t,0)-\bar{F}(t+x,t,t)-\bar{F}(t+x,t+x,0)+\bar{F}(t+x,t+x,t)]}{3[\bar{F}(t,t,0)-\bar{F}(t,t,t)]} \\
&+ \frac{3[\bar{F}(t+x,t+x,0)-\bar{F}(t+x,t+x,t)]}{3[\bar{F}(t,t,0)-\bar{F}(t,t,t)]} \\
&= \frac{6[\bar{F}(t+x,t,0)-\bar{F}(t+x,t,t)]-3[\bar{F}(t+x,t+x,0)-\bar{F}(t+x,t+x,t)]}{3[\bar{F}(t,t,0)-\bar{F}(t,t,t)]}.
\end{aligned}
$$

After algebraic simplifications, it can be shown that

$$P(T_2-t>x|T_2>t,N_2(t)=1)=2e^{-x(1+\theta)}-e^{-x(2+\theta)}.$$

As $\lambda>\theta$, it is easy to verify that

$$e^{-\lambda x-2x}\leq 2e^{-x(1+\theta)}-e^{-x(2+\theta)}.$$

Therefore again we see that the stochastic ordering between these two used systems with EXC components holds true.

3.1 A proposed dynamic signature

Let $T=\phi(X_1,\ldots,X_n)$, $T>t$ and $N(t)=r$. Also suppose $\mathbf{w}_{(n-r)}=(w_{r+1},\ldots,w_n)$ is the vector of dynamic signature defined in Equation (2) and $\mathbf{w}^{(n)}$ is its corresponding usual equivalent dynamic signature which is of order n. Here a simple modified dynamic signature $\mathbf{w}^*$ for $\mathbf{w}^{(n)}$ is proposed that have some useful properties. For $i=1,\ldots,n$ its coordinates are given below

$$w_i^*=\begin{cases} 0 & i\leq r \\ w_i & i>r. \end{cases} \tag{8}$$

In EXC case and in view of (7), it is obvious that

$$P(T-t>x|T>t,N(t)=r)=\sum_{k=1}^{n} w_k^* P(X_{k:n}-t>x|N(t)=r). \tag{9}$$

Theorem 2. *The results of Theorem 1 hold true if $\mathbf{w}_1^{(n)}$ and $\mathbf{w}_2^{(n)}$, are replaced by $\mathbf{w}_1^*$ and $\mathbf{w}_2^*$, respectively and also $(X_{i:n}-t|N_1(t)=r_1)\leq_{st}(Y_{i:n}-t|N_2(t)=r_2)$ for $i=min(r_1,r_2)+1,\ldots,n$.*

Proof. From Equation (9) and in view of the first equality of Equation (7) we have

$$\begin{aligned}
Eg(X) &= \sum_{1}^{n} w^*_{1,k} P(X_{k:n} - t > x | N_1(t) = r_1) = P(T_1 - t > x | T_1 > t, N_1(t) = r_1) \\
&\leq Eg(Y) = \sum_{k=1}^{n} w^*_{2,k} P(X_{k:n} - t > x | N_1(t) = r_1) \\
&= \sum_{k=r_1+1}^{n} w^*_{2,k} P(X_{k:n} - t > x | N_1(t) = r_1) \\
&= \sum_{k=\min(r_1,r_2)+1}^{n} w^*_{2,k} P(X_{k:n} - t > x | N_1(t) = r_1) \\
&\leq \sum_{k=\min(r_1,r_2)+1}^{n} w^*_{2,k} P(Y_{k:n} - t > x | N_2(t) = r_2) \\
&= \sum_{k=r_2+1}^{n} w^*_{2,k} P(Y_{k:n} - t > x | N_2(t) = r_2) = P(T_2 - t > x | T_2 > t, N_2(t) = r_2).
\end{aligned}$$

We note that $P(X_{i:n} - t > x | N(t) = r) = 0$, for $i = 1, \ldots, r$. This completes the proof of the theorem. □

Remark 3. *In comparisons of used systems, Theorem 1 is an extension of part (a) of Theorem 2.6 in Samaniego et al. (2009). On the other hand the conditions of Theorem 2 are less and easier than those of Theorem 1. In Theorem 2, we do not need to compute the usual equivalent dynamic signatures* $\mathbf{w}_1^{(n)}$ *and* $\mathbf{w}_2^{(n)}$*, which are usually obtained after lengthy computations on dynamic signatures. Also the number of stochastic comparisons of conditional random variables in Theorem 2 is* $n - \min(r_1, r_2)$ *which is less than that of Theorem 3.1, that is* n*. One may verify the possibility of happening the following situation:* $\mathbf{w}_1^{(n)} \leq_{st} \mathbf{w}_2^{(n)}$ *but* $\mathbf{w}_1^*$ *and* $\mathbf{w}_2^*$ *are not ordered.*
Obviously, in this case the proposed dynamic signature $\mathbf{w}^*$ *is not applicable and one should use* $\mathbf{w}^{(n)}$*. But the converse of this case may also happen (see the following example when the system components lifetimes are IID).*

Therefore as the above discussed, among $\mathbf{w}^{(n)}$ and $\mathbf{w}^*$ one should clearly first take $\mathbf{w}^*$ into consideration, particularly when $\mathbf{w}^{(n)}$ is not usable. See the following example.

Example 4. Let $\mathbf{s}_1 = (0,0,0,4/5,1/5)$ and $\mathbf{s}_2 = (0,0,3/5,1/5,1/5)$ be the signatures of two systems consisting of different IID components and suppose $N_1(t) = 2$ and $N_2(t) = 3$. From Equation (2), it implies that $\mathbf{w}_1 = (0,4/5,1/5)$ and $\mathbf{w}_2 = (1/2,1/2)$. Also it is easy to see that the equivalent vectors of $\mathbf{w}_1$ and $\mathbf{w}_2$ are $\mathbf{w}_1^{(5)} = 0.01 \times (0,24,34,30,12)$ and $\mathbf{w}_2^{(5)} = (1/5,1/5,1/5,1/5,1/5)$, respectively. Note that in view of the usual stochastic order, $\mathbf{w}_1^{(5)}$ and $\mathbf{w}_2^{(5)}$ are not ordered. Therefore, regardless the system components be common or not, for comparing of two mentioned used systems, part (a) of Theorem 2.6 in Samaniego et al. (2009), and Theorem 1, both are not usable. Whereas,

$$\mathbf{w}_1^* = (0,0,0,4/5,1/5) \leq_{st} \mathbf{w}_2^* = (0,0,0,1/2,1/2).$$

Hence, under the assumptions of Theorem 2, we have $(T_1 - t | T_1 > t, N_1(t) = 2) \leq_{st} (T_2 - t | T_2 > t, N_2(t) = 3)$.

4 Some unique aging properties of the series systems

The preservation of reliability aging classes under the formation of coherent systems is an important topic in reliability studies. This section considers the preservation of IFR and DFR classes in series systems and

particularly shows that among the coherent systems with INID or IID components, the series systems are only DFR closed systems. This result is mainly obtained based on the system signatures. For a detailed study on the preservation of reliability aging classes under the formation of coherent systems refer to Navarro et al. (2013).
Let $T = \phi(X_1, \ldots, X_n)$ be the lifetime of a coherent system in which the component lifetimes X_i's are IID and have a common absolutely continuous distribution F.

Definition 3. *(Samaniego (1985)) A system is said to be IFR (DFR) closed if T has an IFR (DFR) distribution whenever F is an IFR (DFR) distribution. A class of systems is IFR (DFR) closed if each system in the class is IFR (DFR) closed.*

Samaniego (1985) showed that a coherent system with lifetime $T = \phi(X_1, \ldots, X_n)$ and with IID components is IFR closed if and only if

$$k(x) = \frac{\sum_{i=0}^{n-1}(n-i)s_{i+1}\binom{n}{i}x^i}{\sum_{i=0}^{n-1}(\sum_{j=i+1}^{n} s_j)\binom{n}{i}x^i}, \quad x \in (0, \infty)$$

is increasing in x where $\mathbf{s} = (s_1, \ldots, s_n)$ is the system signature.

It is well known that the class of k-out-of-n systems (i.e., systems that work when at least k of their n components work) is IFR closed (see for example Barlow and Proschan (1975)).
Note that all coherent systems with IID components are not necessary IFR closed. For example, in the system with lifetime $T = \max\{X_1, \min(X_2, X_3)\}$, $k(x)$ is not a monotone function.

Remark 4. *It is easy to see that the system is DFR closed if and only if $k(x)$ is decreasing in x. Also note that if a system be IFR and DFR closed both, then $k(x)$ should be a constant function. That is, $s_i = 0$ for $i = 2, \ldots, n$ and therefore $s_1 = 1$. It means that, only the series systems are both IFR and DFR closed. This result is also proved in Navarro et al. (2013). Note that only in a series system $\mathbf{s} = (1, 0, \ldots, 0)$.*

Remark 5. *Using the signatures of all 28 coherent systems of order $n \leq 4$ and verifying the monotonicity of $k(x)$, we observed that in each case it is not decreasing except for series systems in which obviously we have $k(x) = n$. Based on this observation, we conjectured that the series systems are only DFR closed systems. This unique property of the series systems is proved in the following lemma, for both IID and INID cases. First see an example given in Barlow and Proschan (1975) which is needed in the sequel.*

Example 5. Let $T = \max(X_1, X_2)$ where X_1 and X_2 are independent and X_i is exponentially distributed with parameter λ_i, $i = 1, 2$. Then the failure rate function of T is not monotone if $\lambda_1 \neq \lambda_2$, but T has an IFR distribution if $\lambda_1 = \lambda_2$.

Lemma 2. *A coherent system with INID components is a DFR closed system if and only if it be series.*

Proof. It is well-known that if $T = \min(X_1, \ldots, X_n)$ and X_i's are independent with $h_{X_i}(t)$ as the failure rate function of X_i then $h_T(t) = \sum_1^n h_{X_i}(t)$ and therefore a series system is obviously both IFR and DFR closed system. Now to prove the Lemma, we use induction on the values of n. For $n = 2$ and from Example 5, note that T can not be $\max(X_1, X_2)$, as its failure rate function is not monotone and therefore $T = \min(X_1, X_2)$. Now assume that the result of the Lemma holds true for the systems of order $n-1$ that is if $\phi(X_1, \ldots, X_{n-1})$ is DFR when $X_1, \ldots, X_{n-1}$ are DFR then $\phi(X_1, \ldots, X_{n-1}) = \min(X_1, \ldots, X_{n-1})$ or equivalently its structure function is $\phi(x_1, \ldots, x_{n-1}) = \prod_1^{n-1} x_i$.
Now suppose $\phi(X_1, \ldots, X_n)$ is DFR when $X_1, \ldots, X_n$ are DFR. One can consider the system of order n as a system of order 2 in which the subsystem consisting of $x_1, \ldots x_{n-1}$ is its first component and x_n is the second. We note that the main system of order n is DFR closed, therefore both subsystems $\phi(x_1, \ldots, x_{n-1}, 1)$ and $\phi(x_1, \ldots, x_{n-1}, 0)$, which are of order $n-1$ should be DFR closed (otherwise one can easily give the examples of some structures $\phi(X_1, \ldots, X_n)$, such that when the substructures $\phi(X_1, \ldots, X_{n-1}, 1)$ and

$\phi(X_1,\ldots,X_{n-1},0)$ are not DFR closed then the main structure $\phi(X_1,\ldots,X_n)$ is also not DFR closed). Now by pivoting on component n and using induction assumption, we conclude that, both subsystems $\phi(x_1,\ldots,x_{n-1},1)$ and $\phi(x_1,\ldots,x_{n-1},0)$, which are of order $n-1$, have the series structures. On the other hand, in view of the case where $n=2$, these two subsystems and component n are also in series. That is $\phi(x_1,\ldots,x_{n-1},1)=\prod_1^{n-1}x_i\times 1$ and $\phi(x_1,\ldots,x_{n-1},0)=\prod_1^{n-1}x_i\times 0$. Therefore we have

$$\phi(x_1,\ldots,x_n)=x_n\phi(x_1,\ldots,x_{n-1},1)+(1-x_n)\phi(x_1,\ldots,x_{n-1},0)=x_n\times\prod_1^{n-1}x_i+0=\prod_1^{n}x_i.$$

This completes the proof of the lemma. □

Note that Lemma 2 holds true in IID case. That is the series systems with IID components are only DFR closed systems. Therefore Lemma 2 extends a previous result in Lindqvist and Samaniego (2019) for IID case.

Remark 6. *Based on Lemma 2, one can similarly show that the series systems with INID components are only IFR closed systems. Note that unlike the unique property of DFR closedness of the series systems in both IID and INID cases, the class of IFR closed systems with IID components is not contained only by the series systems, as mentioned before it also includes the k-out-of-n systems.*

The result of the Lemma 2 and the mentioned property in Remark 6, extend the Proposition 2.1 in Navarro et al. (2013) from IID to INID case. Proposition 2.5 in Navarro et al. (2013), gives a sufficient condition for a coherent system with INID components to be IFR(DFR) closed. Therefore in view of Remark 6, Lemma 2 presents a more stronger result than the Proposition 2.5 in Navarro et al. (2013). Also for the systems with dependent components, some conditions are obtained in Navarro et al. (2013). They showed that for different coherent systems, some aging classes are preserved.

5 Concluding remarks

This paper presents new properties and applications of system signatures and dynamic signatures, which are essential tools for analyzing the stochastic behavior and aging characteristics of coherent and used systems. Following an extensive review in Section 2, we explored comparisons among coherent systems and examined the relationships between the structure function, reliability function, and system signature. The main contributions are presented in Sections 3 and 4. In Section 3, we investigate comparisons of the residual lifetimes of coherent used systems. By introducing the concept of an equivalent dynamic signature, we derive extended results for systems composed of different exchangeable components. Additionally, we propose an alternative to the equivalent dynamic signature, which exhibits several useful properties for comparing used systems.

Section 4 focuses on unique aging properties of series systems. Specifically, we study the relationship between system signatures and the concepts of IFR and DFR closed systems. It is shown that, under INID and IID assumptions, series systems are only DFR closed, thereby extending existing results in the literature. Illustrative examples are provided throughout the paper to clarify and support the theoretical findings.

Acknowledgement

The authors would like to thanks referees for giving valuable and constructive comments, which led to a significantly improvement in the revised version of the paper.

References

Barlow, R.E. and Proschan F. (1975). *Statistical Theory of Reliability and Life Testing*, Holt, Rinehart and Winston.

Boland, P. (2001). Signatures of indirect majority systems. *Journal of Applied Probability*, **38**, 597–603.

Boland, P. J., and Samaniego, F. J. (2004). The signature of a coherent system and its applications in reliability. In *Mathematical Reliability*: An expository perspective (pp. 3-30). Springer, Boston.

Burkschat, M. and Navarro, J. (2013). Dynamic signatures of coherent systems based on sequential order statistics. *Journal of Applied Probability*, **50**, 272–287.

Burkschat, M. and Samaniego, F. J. (2018). Dynamic IFR concepts for coherent systems. *Statistics & Probability Letters*, **142**, 1–7.

Chahkandi, M., Ruggeri, F., and Suárez-Llorens, A. (2015). A generalized signature of repairable coherent systems. *IEEE Transactions on Reliability*, **65**, 434–445.

Kochar, S., Mukerjee, H., and Samaniego, F. J. (1999). The "signature" of a coherent system and its application to comparisons among systems. *Naval Research Logistics*, **46**, 507–523.

Lindqvist, B. H., and Samaniego, F. J. (2019). Some new results on the preservation of the NBUE and NWUE aging classes under the formation of coherent systems. *Naval Research Logistics*, **66**, 430–438.

Lindqvist, B. H., Samaniego, F. J., and Huseby, A. B. (2016). On the equivalence of systems of different sizes, with applications to system comparisons. *Advances in Applied Probability*, **48**, 332–348.

Navarro, J. (2018). Stochastic comparisons of coherent systems. *Metrika*, **81**, 465–482.

Navarro, J., Balakrishnan, N., and Samaniego, F. J. (2008). Mixture representations of residual lifetimes of used systems. *Journal of applied Probability*, **45**, 1097–1112.

Navarro, J., del Águila, Y., Sordo, M. A., and Suárez-Llorens, A. (2013). Preservation of reliability classes under the formation of coherent systems. *Applied Stochastic Models in Business and Industry*, **30**, 444–454.

Navarro, J., Ruiz, J. M., and Sandoval, C. J. (2005). A note on comparisons among coherent systems with dependent components using signatures. *Statistics and Probability Letters*, **72**, 179–185.

Navarro, J., Samaniego, F. J., and Balakrishnan, N. (2011). Signature-based representations for the reliability of systems with heterogeneous components. *Journal of Applied Probability*, 4**8**, 856–867.

Naqvi, S., Chan, P. S., and Mishra, D. B. (2022). System signatures: A review and bibliometric analysis. *Communications in Statistics-Theory and Methods*, **51**, 1993–2008.

Navarro, J. and Rubio, R. (2011). A note on necessary and sufficient conditions for ordering properties of coherent systems with exchangeable components. *Naval Research Logistics*, **58**, 478–489.

Mahmoudi, M. and Asadi, M. (2011). The dynamic signature of coherent systems. *IEEE Transactions on Reliability*, **60**, 817–822.

Marichal, J. L., Mathonet, P., and Waldhauser, T. (2011). On signature-based expressions of system reliability. *Journal of Multivariate Analysis*, **102**, 1410–1416.

Samaniego, F. J. (2007). *System Signatures and their Applications in Engineering Reliability*. Springer, Boston.

Rao, T. V., and Naqvi, S. (2023). Comparisons of coherent systems with two types of heterogeneous components having proportional reversed hazard rates. *Applied Stochastic Models in Business and Industry*, **40**, 483–511.

Rychlik, T., Navarro, J. and Rubio, R. (2018). Effective procedure of verifying stochastic ordering of system lifetimes. *Journal of Applied Probability*, **55**, 1261–1271.

Sadegh, M. K. (2006). Dynamic signature of a coherent system with exchangeable components. *In: Proceeding of the 13th Iranian statistical conference*, Kerman.

Samaniego, F. J. (1985) On closure of the IFR class under formation of coherent systems. *IEEE Transactions on Reliability*, **34**, 69–72.

Samaniego, F. J. (2006). On the comparison of engineered systems of different sizes. In *Proceedings of the 12th Annual Army Conference on Applied Statistics*, Aberdeen Proving Ground, Army Research Laboratory.

Samaniego, F. J., Balakrishnan, N. and Navarro, J. (2009). Dynamic signatures and their use in comparing the reliability of new and used systems. *Naval Research Logistics*, **56**, 577–591.

Shaked, M. and Shanthikumar, J. G. (2007). *Stochastic Orders*, Springer, New York.

Stochastic Models in Probability and Statistics
Vol. 2, No. 2, 2025, pp. 133–157. Research Paper

Higher-order moments of Markov switching bilinear models: theory and empirical evidence

Abdelouahab Bibi*

Department of Mathematics, University Fréres Mentouri Constantine 1, Constantine, Algeria
Email(s): abd.bibi@gmail.com

Abstract. In this paper, we consider a *Markov-switching bilinear process* (*MS-BL*) that exhibits rich dynamic behavior and plays an important role in modeling non-Gaussian data characterized by structural breaks. In such models, the parameters depend on an unobservable (hidden) Markov chain with a finite state space. Although numerous recent studies have focused on the statistical aspects of Markov-switching models, systematic investigations of the probabilistic properties of this class of nonlinear models remain relatively scarce. So, we derive conditions for stationarity and compute the moments of the process up to the third order. Our analysis reveals that the conditions ensuring local stationarity within each regime of the observed process are neither sufficient nor necessary. Furthermore, we show that the second-order structure of the process is analogous to that of a Markov-switching *ARMA* (*MS-ARMA*) model with an additional uncorrelated white noise component. Therefore, the examination of higher-order moments becomes essential to distinguish between (locally) linear and nonlinear models. To illustrate the practical relevance of our theoretical results, we conduct Monte Carlo simulation studies and apply the proposed model to the exchange rate of the Algerian Dinar against the Euro. The empirical findings indicate that the proposed approach provides a better fit and demonstrates superior performance compared to alternative models.

Keywords: Higher-order moments; Markov-switching bilinear models; Stationarity.

1 Motivations

Markov-switching (*MS*) time series models proposed by Hamilton (1989) have recently received a growing interest in several areas of statistics because of their ability to describe adequately various financial time series and continue to gain more popularity, especially to model empirical macroeconomics and dynamic econometrics time series $(X_t)_{t\in\mathbb{Z}}$, $\mathbb{Z} = \{0, \pm 1, \pm 2, ...\}$. The advantages of *MS* models are multiple, for instance, among others:
(i) *MS* models are nonlinear models (even locally) because linear models are not, in general, always suitable for use.
(ii) Higher flexibility in capturing the persistence and/or asymmetric effect in datasets.

*Corresponding author

Received: 10 July 2025 / Accepted: 29 October 2025
DOI: 10.22067/smps.2025.94345.1049

(iii) In modeling time series which exhibit smooth or abrupt structure which occur frequently or occasionally depending on the transition probability of the chain.
So, there exist various different ways to model such a series exhibiting break changes through a finite number of "regimes". A rather general model is given by

$$X_t = f_{s_t}\left(X_{t-i}, e_{t-j}, 0 < i \leq P, 0 < j \leq Q\right) + e_t,$$

for some measurable function $f_.$ depending on some finite state Markov chain $(s_t)_t$ that control the change of regimes and some innovation process $(e_t)_t$. However, in literature, some locally (given s_t) linear or nonlinear models (determined by $f_.$) were investigated in order to study the probabilistic and statistical properties of such models. Indeed, the stationary $ARMA$ models in which the parameters are allowed to change through time according to Markov chain denoted by $MS-ARMA$ have, got considerable attention recently. For instance, Yang (2000), Francq and Zakoïan (2001), Stelzer (2009), Lee (2005), Yao and Attali (2000), Boubacar and Rabehasaina (2020) and the references therein, are references aims to describe the probabilistic and/or statistical properties of $MS-ARMA$ models. Namely, conditions assuring the strict stationarity, ergodicity, the existence of higher-order moments and spectral representation. On the other hand, to develop some appropriate statistical methods and their asymptotic inference. Francq and Zakoïan (2005), Bauwen et al. (2010), Alemohammad et al. (2020), Bibi and Ghezal (2015) and Wee et al. (2022) have proposed a $MS-GARCH$ model. So, the probabilistic structure, as in $MS-ARMA$ models were investigated and some procedures for estimating and forecasting the MS_GARCH model was studied. Recently, Bibi and Ghezal (2015) and Bibi and Hamdi (2025) and have introduced a new class of $MS-$bilinear ($MS-BL$ for short) model where several probabilistic properties were studied and explicit conditions ensuring the existence of a strictly stationary solution of such a model to belong in $\mathbb{L}_2$ are given. Moreover, it is also known that some bilinear processes have properties that are similar to those of an autoregressive conditionally heteroscedastic ($ARCH$) model, which plays important role in financial mathematics see subsection 3.1.
In this paper, a process $(X_t)_t$ defined on some probability space $(\Omega, \Im, P)$ is called $MS-BL(p,q,P,Q)$ if it generated by the following stochastic difference equation:

$$X_t = a_0(s_t) + \sum_{i=1}^{p} a_i(s_t)X_{t-i} + \sum_{j=0}^{q} b_j(s_t)e_{t-j} + \sum_{j=1}^{Q}\sum_{i=1}^{P} c_{ij}(s_t)X_{t-i}e_{t-j}. \tag{1}$$

In (1), the functions $a_i(s_t)$, $b_j(s_t)$ and $c_{ij}(s_t)$ depend upon a Markov chain $(s_t)_t$ that controls the dynamics of X_t and is subject to the following assumption:

The Markov chain $(s_t)_t$ is irreducible and aperiodic with a finite state space $\mathbb{S} = \{1, ..., d\}$, the $n-$step transition probability matrix, that determines the evolution in $\mathbb{S}$ is given by

$$P^{(n)} = \left(p_{ij}^{(n)}\right)'_{(i,j)\in\mathbb{S}\times\mathbb{S}},$$

where $p_{ij}^{(n)} = P(s_t = j|s_{t-n} = i)$, so by Chapman-Kolmogorov Equations $P^{(n)} = P^n$. The one-step transition probability matrix $P := (p_{ij})$ where $p_{ij} := p_{ij}^{(1)} = P(s_t = j|s_{t-1} = i)$ for $i,j \in \mathbb{S}$, such that $\sum\limits_{j=1}^{d} p_{ij} = 1$, for all i. The stationary distribution of the Markov chain $(s_t)_t$ will be denoted by $\underline{\Pi} = (\pi(1), ..., \pi(d))'$ that solves the equation $P'\underline{\Pi} = \underline{\Pi}$ where $\pi(i) = P(s_0 = i)$, $i = 1, ..., d$. The chain $(s_t)_t$ is said to be independent if $p_{ij} = \pi(j)$ for all $i \in \mathbb{S}$.

The innovation process $(e_t)_t$ is assumed to be independent and identically distributed ($i.i.d$) with mean 0 and variance σ_2. In addition, we shall assume that e_t and $\{(X_{s-1}, s_t), s \leq t\}$ are independent.

As already pointed out by Bibi and Ghezal (2015), the $MS-BL(p,q,P,Q)$ includes as special cases several classes of interesting models having been investigated in the literature. It is worth noting that the key difference between $MS-BL$ and a threshold model is that the former assumes that the underlying state process that gives rise to the nonlinear dynamics (regime switching) is latent, whereas threshold models commonly allow the nonlinear effect to be driven by observable variables but assume the number of thresholds and the threshold values to be unknown. Before we proceed, we will first introduce some algebraic notations and definitions.

1.1 Algebraic notations and definitions

Some notations are used throughout the paper:

- For some specifications of the chain $(s_t)_t$, i.e., constant $(d=1)$, independent and dependent chain, in the sequel we shall indicate the corresponding models by $C-BL$, $I-BL$ and $D-BL$ respectively.
- $I_{(n)}$ is the $n\times n$ identity matrix, $\mathbb{I} := \left(I_{(s)} \vdots \ldots \vdots I_{(s)}\right)_{s\times ds}$, is the $d-$block matrix, and $\mathbf{1}_{(d)}$ (resp. $\underline{\mathbf{1}}_{(d)}$) is a $d\times d$ matrix (resp. $d\times 1$ vector) whose components are matrix unity (resp. 1).
- $O_{(k,l)}$ denotes the matrix of order $k\times l$ whose entries are zeros, for simplicity we set $O_{(k)} := O_{(k,k)}$ and $\underline{O}_{(k)} := O_{(k,1)}$.
- The spectral radius of a square matrix M is denoted by $\rho(M)$, $\|.\|$ denotes any operator norm on the set of $m\times n$ and $m\times 1$ matrices, $\otimes$ is the usual Kronecker product of matrices and $M^{\otimes r} = M\otimes M\otimes ...\otimes M$, $r-times$. If $(M(i), i\in I)$ is $n\times n$ matrices sequence, we shall note for any integer l and j, $\prod\limits_{i=l}^{j} M(i) = M(l)M(l+1)\ldots M(j)$ if $l\leq j$ and $I_{(n)}$ otherwise.
- For any function $f:\mathbb{S}\to\mathscr{M}_{n\times m}(\mathbb{R})$, where $\mathscr{M}_{n\times m}(\mathbb{R})$ denotes the space of real $n\times m$ matrices, we write

$$\mathbb{P}(f^{\otimes r}) = \begin{pmatrix} p_{11}f^{\otimes r}(1).. & p_{d1}f^{\otimes r}(1) \\ \vdots & \vdots \\ p_{1d}f^{\otimes r}(d).. & p_{dd}f^{\otimes r}(d) \end{pmatrix},\ \mathbb{P}^{(n)}(f^{\otimes r}) = \begin{pmatrix} p^{(n)}_{11}f^{\otimes r}(1).. & p^{(n)}_{d1}f^{\otimes r}(1) \\ \vdots & \vdots \\ p^{(n)}_{1d}f^{\otimes r}(d).. & p^{(n)}_{dd}f^{\otimes r}(d) \end{pmatrix}$$

  and

$$\underline{\Pi}(f^{\otimes r}) = \begin{pmatrix} \pi(1)f^{\otimes r}(1) \\ \vdots \\ \pi(d)f^{\otimes r}(d) \end{pmatrix}.$$

  In the sequel, we will use the following result due to Francq and Zakoïan (2005) stated in the next lemma.

Lemma 1. *1. For $i\geq 1$, if $(Z_{t-i})_t$ is an integrable random variable belonging to $\sigma(e_{t-s}, s\geq i)$, then*

$$\pi(k)E\{Z_{t-i}|s_t = k\} = \sum_{j=1}^{d} E\{Z_{t-i}|s_{t-i} = j\}\, p^{(i)}_{jk}\pi(j).$$

2. If $f:\mathbb{S}\to\mathscr{M}_{n\times n}(\mathbb{R})$ and $g:\mathbb{S}\to\mathscr{M}_{n\times 1}(\mathbb{R})$, then for any $k>0$ and $\tau>k$,

$$E\left\{f(s_t)f(s_{t-1})...f(s_{t-k+1})\underline{g}(s_{t-k})|s_{t-k}\right\} = \mathbb{I}\{\mathbb{P}(f)\}^k\,\underline{\Pi}(\underline{g}).$$

Remark 1. *[Independent switching]. For the model $I-BL$, then $(\mathbb{P}(f))^k = \mathbb{P}\left(f(.)\left(E\{f(s_t)\}\right)^{k-1}\right)$ and $\mathbb{IP}(f) = [E\{f(s_t)\}, ..., E\{f(s_t)\}]$ where in the last equality, the matrix $E\{f(s_t)\}$ is duplicated $d-$times.*

We arrange the rest of the paper in the following manner. In the next section, we give the Markovian state-space representation, which is used to derive conditions for stationarity (in the strong and weak sense) and a recursive formula for the higher-order moments. Section 3 is devoted to giving explicit expressions of the second, third and fourth-order moments of some particular $MS-BL$ models. Section 4 provides results of a Monte Carlo experiment and section 5 concludes the paper.

2 Markovian representation of $MS-BL$ and its properties

In the rest of the paper, we shall restrict ourselves to the case when $b_j(.) = 0$, $j = 1,...,q$ in (1), i.e., without moving average part

$$X_t = \sum_{i=1}^{p} a_i(s_t)X_{t-i} + \sum_{j=1}^{Q}\sum_{i=1}^{p} c_{ij}(s_t)X_{t-i}e_{t-j} + e_t, \tag{2}$$

which denoted also hereafter $MS-BL(p,0,p,Q)$. Because it is difficult to handle the product terms like X_te_{t-j}, $j > 0$, Liu (1992) when $d = 1$, introduced the so-called lower triangular model for which $c_{ij}(.) = 0$, if $i < j$. So in this paper we extend the lower tridiagonal model to include a $MS-BL$ one, i.e.,

$$X_t = \sum_{i=1}^{p} a_i(s_t)X_{t-i} + \sum_{j=1}^{Q}\sum_{i=j}^{p} c_{ij}(s_t)X_{t-i}e_{t-j} + e_t. \tag{3}$$

The representation (3) has however the advantage to admit the state-space representation $X_t = \underline{H}'\underline{Y}_t$ where

$$\underline{Y}_t = \Gamma_{s_t}(e_t)\underline{Y}_{t-1} + \underline{\eta}(e_t), \tag{4}$$

with $\Gamma_{s_t}(e_t) = \Gamma_0(s_t) + e_t\Gamma_1(s_t)$ where the matrices $\Gamma_0(s_t)$, $\Gamma_1(s_t)$ and the vectors $\underline{\eta}(e_t)$, $\underline{H}$ are explicitly given in Bibi and Ghezal (2015). Note that the representation (4) shows that the $MS-BL(p,0,p,Q)$ can be represented as a multidimensional first-order random coefficient Markov-switching Autoregressive ($MS-RCAR(1)$) model, and hence the extended process $\underline{Z}_t := (\underline{Y}_t', s_t)'$, is an aperiodic Markov chain on $(\mathscr{X}, \mathscr{B}(\mathscr{X}))$, where $\mathscr{X} = \mathbb{R}^s \times \mathbb{S}$. This compact representation allows us to provide a necessary and sufficient condition for the existence of strict stationery, ergodic, and $\Im_t = \sigma(e_{t-n}, s_{t-n}, n \geq 0)$ −measurable (or causal) solutions for (3). These concepts are ensured under the strict negativeness of the Lyapunove exponent defined by $\gamma(\Gamma) := \inf_{t\geq 1} \frac{1}{t} E\left\{\log\left\|\prod_{i=0}^{t-1}\Gamma_{s_{t-i}}(e_{t-i})\right\|\right\}$ (the chosen of the norm is unimportant in this definition). However, it follows from Bibi and Ghezal (2015) that the unique strictly stationary and ergodic solution of (4) is given by the first component of

$$\underline{Y}_t = \sum_{k=0}^{\infty}\left\{\prod_{i=0}^{k-1}\Gamma_{s_{t-i}}(e_{t-i})\right\}\underline{\eta}(e_{t-k}), \text{ a.s.,} \tag{5}$$

with the usual convention $\prod_{i=k}^{k'} = 1$, whenever $k' < k$.

Remark 2. *For the process $(X_t)_t$ defined by $X_t = c_{11}(s_t)X_{t-1}e_{t-1} + e_t$, the Lyapunov exponent is*

$$\sum_{i=1}^{d}\pi(i)E\{\log(|c_{11}(i)e_0|)\},$$

so $\prod_{i=1}^{d}|c_{11}(i)|^{\pi(i)} < e^{-E\{\log|e_0|\}}$ constitutes the necessary and sufficient condition for strict stationarity.and ergodic solution. Note that when $(e_t)_t$ is Gaussian, the necessary and sufficient condition reduces to

$\prod_{i=1}^{d}|c_{11}(i)|^{\pi(i)} < 1.88736$. *It is worth noting that this example shows the existence of explosive regimes (i.e. ,*$|c_{11}(i)| > 1$ *compared with the standard case) does not preclude the strict stationarity.*

Recalling here (interested readers are advised to see Bibi and Ghezal (2015) for more details) that if $\sigma_{2m+1} = E\left\{e_t^{2m+1}\right\} = 0$, $\sigma_{2m} = E\left\{e_t^{2m}\right\} < +\infty$ for any integer m, and if

$$\lambda_{(m)} := \rho\left(\mathbb{P}\left(\Gamma^{\otimes m}\right)\right) < 1, \tag{6}$$

where $\Gamma^{\otimes m} := (\Gamma^{\otimes m}(i), 1 \leq i \leq d)$ with $\Gamma^{\otimes m}(.) = E\left\{\Gamma_{s_t}^{\otimes m}(e_t)|s_t = .\right\}$, then equation (4) has a unique strictly stationary solution $(\underline{Y}_t)_t$ given by (5) and the process $(\underline{H}'\underline{Y}_t)_t$ is also a unique, strictly stationary, causal and ergodic solution of equation (3) which satisfies $X_t^{2m} \in \mathbb{L}_{2m}$.

Remark 3. *For the model $I-BL$ the condition for the existence a strictly stationary solution in $\mathbb{L}_m$ is that $\lambda_{(m)} := \rho\left(E\left\{\Gamma_{s_t}^{\otimes m}\right\}\right) < 1$ where $E\left\{\Gamma_{s_t}^{\otimes m}\right\} = \sum_{k=1}^{d}\Gamma^{\otimes m}(k)\pi(k)$.*

2.1 Computation of the higher-order moments

Once the second-order stationarity condition is established, it is useful to compute the expectation and some cumulants of the process $(\underline{Y}_t)_t$. A property which will be heavily used in the sequel, associated with representation (4) is given in the following lemma.

Lemma 2. *Consider the representation (4), then for any integer $m \geq 0$, we have*

$$\underline{Y}_t^{\otimes m} = \sum_{i=0}^{m}\Psi_i^{(m)}(s_t, e_t)\,\underline{Y}_{t-1}^{\otimes i}, \tag{7}$$

where the matrices $\Psi_i^{(m)}(s_t, e_t)'s$ are uniquely determined by $\Gamma_{s_t}(e_t)$ and $\underline{\eta}(e_t)$ according to the following recursions

$$\Psi_0^{(m)}(s_t, e_t) = \underline{\eta}^{\otimes m}(e_t),\ \Psi_1^{(1)}(s_t, e_t) = \Gamma_{s_t}(e_t), \quad \text{and for any } m > 0,$$
$$\Psi_i^{(m+1)}(s_t, e_t) = \underline{\eta}(e_t) \otimes \Psi_i^{(m)}(s_t, e_t) + \Gamma_{s_t}(e_t) \otimes \Psi_{i-1}^{(m)}(s_t, e_t),$$

with the convention $\Psi_i^{(m)}(s_t, e_t) = 0$ when $i > m$ or $i < 0$ and $\underline{Y}_t^{\otimes 0} = \Psi_0^{(0)}(s_t, e_t) = 1$.

Proof. The formula (7) for $m = 1$ is given by (4). Assuming that (7) holds true for some $m \geq 1$, then we have

$$\begin{aligned}
\underline{Y}_t^{\otimes(m+1)} &= \sum_{i=0}^{m}\left(\Gamma_{s_t}(e_t)\underline{Y}_{t-1} + \underline{\eta}(e_t)\right) \otimes \Psi_i^{(m)}(s_t, e_t)\,\underline{Y}_{t-1}^{\otimes i} \\
&= \sum_{i=0}^{m}\left\{\underline{\eta}(e_t) \otimes \left\{\Psi_i^{(m)}(s_t, e_t)\,\underline{Y}_{t-1}^{\otimes i}\right\} + \left(\Gamma_{s_t}(e_t) \otimes \Psi_i^{(m)}(s_t, e_t)\right)\left(\underline{Y}_{t-1} \otimes \underline{Y}_{t-1}^{\otimes i}\right)\right\} \\
&= \sum_{i=0}^{m+1}\left\{\left\{\underline{\eta}(e_t) \otimes \Psi_i^{(m)}(s_t, e_t)\right\}\underline{Y}_{t-1}^{\otimes i} + \left\{\Gamma_{s_t}(e_t) \otimes \Psi_{i-1}^{(m)}(s_t, e_t)\right\}\underline{Y}_{t-1}^{\otimes i}\right\}
\end{aligned}$$

and the result follows. □

Now, we note $\underline{U}^{(m)} = \left(\pi(k)E\left\{\underline{Y}_t'^{\otimes m}|s_t=k\right\}, k=1,..,d\right)'$, then

$$\begin{aligned}\pi(k)E\left\{\underline{Y}_t^{\otimes m}|s_t=k\right\} &= \sum_{i=0}^{m}\Psi_i^{(m)}(k)E\left\{\underline{Y}_{t-1}^{\otimes i}|s_t=k\right\}\pi(k)\\ &= \sum_{i=0}^{m}\Psi_i^{(m)}(k)\sum\nolimits_{j=1}^{d}E\left\{\underline{Y}_{t-1}^{\otimes i}|s_{t-1}=j\right\}p_{jk}\pi(j),\end{aligned} \tag{8}$$

where $\Psi_i^{(m)}(k) = E\left\{\Psi_i^{(m)}(s_t,e_t)|s_t=k\right\}$, so $\underline{U}^{(m)} = \sum\limits_{i=0}^{m}\mathbb{P}\left(\Psi_i^{(m)}(k)\right)\underline{U}^{(i)}$, hence under the condition (6), $E\left\{\underline{Y}_t^{\otimes m}\right\} = \mathbb{I}\underline{U}^{(m)}$ and $\underline{U}^{(m)} = \left(I_{(ds^m)} - \mathbb{P}\left(\Psi_m^{(m)}(k)\right)\right)^{-1}\left\{\sum\limits_{i=0}^{m-1}\mathbb{P}\left(\Psi_i^{(m)}(k)\right)\underline{U}^{(i)}\right\}$.

Example 1. Explicit formulae for some power m are addressed below

$$\Psi_i^{(1)}(k) = \begin{cases}\underline{\eta}^{(0)}, & i=0,\\ \Gamma_0(k), & i=1.\end{cases}$$

$$\Psi_i^{(2)}(k) = \begin{cases}\underline{\eta}^{(0)\otimes 2}, & i=0,\\ \underline{\eta}^{(0)}\otimes\Gamma_0(k)+\underline{\eta}^{(1)}\otimes\Gamma_1(k)+\Gamma_0(k)\otimes\underline{\eta}^{(0)}+\Gamma_1(k)\otimes\underline{\eta}^{(1)}, & i=1,\\ \Gamma_0^{\otimes 2}(k)+\sigma_2\Gamma_1^{\otimes 2}(k), & i=2.\end{cases}$$

For $m=3$, the non-zeros terms in $\Psi_i^{(3)}(k)$, $i=0,1,2$ and 3 are

$$\Psi_i^{(3)}(k) = \begin{cases}\underline{\eta}^{(0)\otimes 3}, & i=0,\\ \underline{\eta}^{(0)\otimes 2}\otimes\Gamma_0 k)+\underline{\eta}^{(1)\otimes 2}\otimes\Gamma_1 k)+\underline{\eta}^{(0)}\otimes\Gamma_0(k)\otimes\underline{\eta}^{(0)}+\underline{\eta}^{(0)}\otimes\Gamma_1(k)\otimes\underline{\eta}^{(1)(0)} & \\ +\Gamma_0(k)\otimes\underline{\eta}^{(0)}+\Gamma_1(k)\otimes\underline{\eta}^{(1)}, & i=1,\\ \underline{\eta}^{(0)}\otimes\Gamma_0^{\otimes 2}(k)+\underline{\eta}^{(2)}\otimes\Gamma_1^{\otimes 2}(k)+\underline{\eta}^{(1)}\otimes(\Gamma_0(k)\otimes\Gamma_1(k)+\Gamma_1(k)\otimes\Gamma_0(k)) & \\ +\Gamma_0(k)\otimes\underline{\eta}^{(0)}\otimes\Gamma_0(k)+\Gamma_0(k)\otimes\underline{\eta}^{(1)}\otimes\Gamma_1(k)+\Gamma_1(k)\otimes\underline{\eta}^{(1)}\otimes\Gamma_0(k) & \\ +\Gamma_1(k)\otimes\underline{\eta}^{(2)}\otimes\Gamma_1(k) & i=2,\\ \left(\Gamma_0^{\otimes 2}(k)+\sigma_2\Gamma_1^{\otimes 2}(k)\right)\otimes\Gamma_0(k)+\sigma_2(\Gamma_0(k)\otimes\Gamma_1(k)+\Gamma_1(k)\otimes\Gamma_0(k))\otimes\Gamma_1(k), & i=3\end{cases}$$

where $\underline{\eta}^{(i)} = E\left\{e_t^i\underline{\eta}(e_t)\right\}$, $i=0,1,....$ The fourth$-$order moments is also useful to study which may be deduced from the recursion $\underline{U}^{(m)}$.

Note here that the recursion $\underline{U}^{(m)}$ demonstrates the dependence of the $m-th$ moment terms on the $(m-1)th-$moment. Provided that the $(m-1)th-$moment.converges, and the summation in (8) is also stable, this term will converge. Moreover, we note that the matrices $\Psi_i^{(m)}(k)$ are not zero for any $k\in\mathbb{S}$.

3 Case studies

In this section we examine the following particular cases of $MS-BL$ models (3), i.e.,

$$\begin{cases}\text{(Superdiagonal model):} & X_t = a(s_t)X_{t-l}e_{t-k}+e_t,\ 0<k<l, & \text{(MS-SBL)}\\ \text{(Diagonal model):} & X_t = a(s_t)X_{t-k}e_{t-k}+e_t,\ 0<k, & \text{(MS-DBL)}\end{cases} \tag{9}$$

where for the sake of generality it is assumed that $(e_t)_t$ is *i.i.d.*, Gaussian sequence, and we set $\sigma_r = E\left\{e_t^r\right\}$. So, the even moments of $(e_t)_t$ up to eight order are $\sigma_2=\sigma^2$, $\sigma_4=3\sigma_2^2$, $\sigma_6=15\sigma_2^3$, and $\sigma_8=105\sigma_2^4$ while all the odd moments are equal to zero. The above models was studied in non switching framework (i.e., $d=1$) by, among others, Gabr (1988) and Martin (1999) who gave some general discussions on the properties of such models including stationarity and moments properties. Moreover, Grange and Anderson (1978) studied the superdiagonal in detail and showed that this series might be mistaken as a white noise. In the following we give an explicit expression of the higher-order moments of the above models. Let us consider

$$\gamma_X(i) = E\{(X_t-\mu)(X_{t-i}-\mu)\} = \mu_X(i)-\mu^2, \tag{10}$$

$$\gamma_X(i,j) = E\left\{(X_t-\mu)(X_{t-i}-\mu)(X_{t-j}-\mu)\right\} = \mu_X(i,j)-\mu\{\mu_X(i)+\mu_X(j)+\mu_X(i-j)\}+2\mu^3, \tag{11}$$

where $\mu=E\{X_t\}$, $\mu_X(i)=E\{X_tX_{t-i}\}$ and $\mu_X(i,j)=E\left\{X_tX_{t-i}X_{t-j}\right\}$. In this illustration, the autocorrelations of $\left(X_t^2\right)$ noted $\rho_2(i) = (\mu_{(2)}(i)-\mu_X^2(0))/Var\left\{X_t^2\right\}$ where $\mu_{(2)}(i)=E\left\{X_t^2X_{t-i}^2\right\}$, it be use as a power criterion for bilinear model identification, which often replaced by a standardized third central moments, i.e.,

$$\begin{aligned}\rho_3(i,j) &= E\left\{(X_t-\mu)(X_{t-i}-\mu)(X_{t-j}-\mu)\right\}/\{Var\{X_t\}\}^{3/2}\\ &= \left\{\mu_X(i,j)-\mu\{\mu_X(j)+\mu_X(i)+\mu_X(i-j)\}+2\mu^3\right\}/\gamma_X^{3/2}(0).\end{aligned}$$

The last moment has several features with respect to $\rho_2(i)$. For instance, the coefficients of skewness of $(X_t)_t$ may be derived from $\rho_3(i,j)$. Moreover, if $\left(X_t^2\right)_t$ is generated by a (locally) linear model then $\rho_2(i)$ can be nonzero whereas $\rho_3(i,j)$ will be zero for all lags i and j. Also $\rho_3(i,j)$ should provide more information about the model under study than the autocorrelation of $\left(X_t^2\right)_t$. Note that from Gabr (1988), the following symmetry relations hold $\gamma_X(i,j)=\gamma_X(j,i)=\gamma_X(-i,j-i)=\gamma_X(i-j,-j)$ where $i,\ j\in\mathbb{Z}$. So, it is sufficient to calculate $\gamma_X(i,j)$ for $0\le i\le j$. In the sequel we shall note for any $i\ge 1$, $a^i=\left(a^i(j), j\in\mathbb{S}\right)$ and $\prod\limits_{j=1}^{i}\mathbb{P}^{(k)}(.)=\mathbb{P}^{i(k)}(.)$.

3.1 Superdiagonal model

Note that the $MS-SBL$ model may be written as $X_t=a(s_t)X_{t-k}e_{t-k+m}+e_t, 2\le k, 1\le m\le k-1$, which is conditionally heteroscedastic (but not a $MS-ARCH$ model). Indeed, for $k=2$, let $h_t(l):=E\left\{\pi(l)X_t^2I_{\{s_t=l\}}|\Im_{t-2}\right\}$, $l\in\mathbb{S}$, then we have

$$\begin{aligned}h_t(l) &= \sigma_2a^2(l)\pi(l)E\left\{X_{t-2}^2I_{\{s_t=l\}}|\Im_{t-2}\right\}+\pi(l)\sigma_2\\ &= \sigma_2a^2(l)\sum_{l'=1}^{d}E\left\{X_{t-2}^2I_{\{s_{t-2}=l'\}}|\Im_{t-2}\right\}P_{ll'}^{(2)}\pi(l')+\pi(l)\sigma_2\\ &= \sigma_2a^2(l)X_{t-2}^2\sum_{l'=1}^{d}I_{\{s_{t-2}=l'\}}P_{ll'}^{(2)}\pi(l')+\pi(l)\sigma_2.\end{aligned}$$

So, $h_t=\sum_{l=1}^{d}h_t(l)$ is given by $h_t=\left\{\underline{\mathbf{1}}\mathbb{P}^{(2)}\left(\sigma_2a^2\right)\underline{Z}_{t-2}\right\}X_{t-2}^2+\sigma_2$ where $\underline{Z}_t=\left(I_{\{s_t=1\}},\ldots,I_{\{s_t=d\}}\right)'$, this show that $MS-SBL$ is an $ARCH$ model. From the discussion in subsection 2.1, it can be shown that a sufficient condition for the strict stationarity solution in $\mathbb{L}_{2m}$ is that $\rho\left(\mathbb{P}^{(k)}\left(\sigma_{2m}a^{2m}\right)\right)<1$, $m\ge 1$.

Lemma 3. [*First and second-order moments*] *For the $MS-SBL$, assume that $\rho\left(\mathbb{P}^{(k)}\left(\sigma_2a^2\right)\right)<1$, which ensures the second-order stationarity, then we have*

1. $\mu_{(1)}(k)=E\{X_t\}=0$, $E\left\{X_te_t^2\right\}=0$, $E\left\{X_t^2e_t\right\}=0$, $E\left\{X_t^2e_t^3\right\}=0$, $E\{X_te_t\}=\sigma_2$, $E\left\{X_te_t^3\right\}=\sigma_4$,
2. $\mu_{(2)}(k)=E\left\{X_t^2\right\}=\underline{\mathbf{1}}\underline{\mu}_{(2)}(k)$ *where* $\underline{\mu}_{(2)}(k)=\left(I_{(d)}-\mathbb{P}^{(k)}\left(\sigma_2a^2\right)\right)^{-1}\underline{\Pi}(\sigma_2)$,

3. $\widetilde{\mu}_{(2)}(k) = E\left\{X_t^2 e_t^2\right\} = \underline{\mathbf{1}}\widetilde{\underline{\mu}}_{(2)}(k)$ *where* $\widetilde{\underline{\mu}}_{(2)}(k) = \sigma_2^2 \mathbb{P}^{(k)}\left(a^2\right)\left(I_{(d)} - \mathbb{P}^{(k)}\left(\sigma_2 a^2\right)\right)^{-1} \underline{\Pi}(\sigma_2) + \underline{\Pi}(\sigma_4)$.
4. $\gamma_X(i) = \mu_{(2)}(k)\,\delta_{\{i=0\}}$.

Proof. The proof is straightforward and hence omitted. □

The nullity of $\gamma(i)$ for $i > 0$, shows that the $MS-SBL$ appears as a sequence of weak white noise ($0-$dependent). So, in order to provide more informations, we need to investigate some moments of order greater than 2.

Lemma 4. *Under the condition of Lemma 3, all the third-order moments $\gamma(i,j)$ of the $MS-SBL$, are equal to zero except $i = k-m$ and $j = k$ for which we have*

$$\gamma_X(k-m,k) = \underline{\mathbf{1}}\mathbb{P}^{(k-m)}(a)\,\mathbb{P}^{(m)}(\sigma_2)\left(I_{(d)} - \mathbb{P}^{(k)}\left(\sigma_2 a^2\right)\right)^{-1} \underline{\Pi}(\sigma_2).$$

Proof. It is easy to see that $\gamma_X(0,0) = 0$. For $i = k-m, j = k$, we have

$$\begin{aligned}\pi(l)\,E\left\{X_t X_{t-k} X_{t-k+m} | s_t = l\right\} &= \pi(l)\,a(l)\,E\{X_{t-k+m} e_{t-k+m} X_{t-k}^2 | s_t = l\} \\ &= \underline{\mathbf{1}}\mathbb{P}^{(k-m)}(a)\,\underline{U},\end{aligned}$$

where

$$\underline{U} = (E\left\{X_{t-k+m} e_{t-k+m} X_{t-k}^2 | s_{t-k+m} = l\right\}\pi(l), l = 1,..,d)' = \mathbb{P}^{(m)}\left(\sigma^2\right) E\{X_t^2 | s_t = u\}\pi(u), u = 1,..,d)'.$$

So,

$$\gamma_X(k-m,k) = \underline{\mathbf{1}}\mathbb{P}^{(k-m)}(a)\,\mathbb{P}^{(m)}(\sigma_2)\left(I_{(d)} - \mathbb{P}^{(k)}\left(\sigma_2 a^2\right)\right)^{-1} \underline{\Pi}(\sigma_2).$$

□

Lemma 5. *For the $MS-SBL$, we have under the condition $\rho\left(\mathbb{P}^{(k)}\left(\sigma_{2l} a^{2l}\right)\right) < 1,\ l \geq 1$.*

1. $\mu_{(r)}(k) = E\{X_t^r\} = \underline{\mathbf{1}}\underline{\mu}_{(r)}(k)\,I_{\{r=2l\}}$ *where*

$$\underline{\mu}_{(2l)}(k) = \left(I_{(d)} - \mathbb{P}^{(k)}\left(\sigma_{2l} a^{2l}\right)\right)^{-1} (\textstyle\sum_{i=1}^{l-1} C_{2l}^{2i} \sigma_{2(l-i)} \mathbb{P}^{(k)}(\ \sigma_{2i} a^{2i})\,\underline{\mu}_{(2i)}(k) + \underline{\Pi}(\sigma_{2l})).$$

2. $\widetilde{\mu}_{(r)}(k) = E\{(X_t e_t)^r\} = \underline{\mathbf{1}}\widetilde{\underline{\mu}}_{(r)}(k)$ *where* $\widetilde{\underline{\mu}}_{(r)}(k) = \sigma_{2r} + \underline{\mathbf{1}}\left(\sum_{i=1}^{\left[\frac{r}{2}\right]} C_r^{2i} \mathbb{P}^{(k)}\left(\sigma_{2r}\ a^{2i}\right)\underline{\mu}_{(2i)}(k)\right)$, $[x]$ *is the integer part of* x.

Proof. 1. For the first assertion, we have for any integer $r \geq 1$,

$$X_t^r = e_t^r + \textstyle\sum_{i=1}^{r} C_r^i a^i(s_t)\,X_{t-k}^i e_{t-k+m}^i e_t^{r-i}.$$

So if $r = 2l+1$, then due to Gaussianity of $(e_t)_t$, the expectation of the first term is 0 and because if i is even integer, $r-i$ is odd and vice versa. Therefore the expectation of the second sum is also 0. In contrast if $r = 2l$ say, then

$$\begin{aligned}E\left\{X_t^{2l} | s_t = j\right\}\pi(j) &= E\left\{e_t^{2l} | s_t = j\right\}\pi(j) + \textstyle\sum_{i=1}^{l} C_{2l}^{2i} a^{2i}(j)\,E\left\{X_{t-k}^{2i} | s_t = j\right\}\pi(j)\,\sigma_{2i}\sigma_{2(l-i)} \\ &= \sigma_{2l}\pi(j) + \textstyle\sum_{i=1}^{l} C_{2l}^{2i} a^{2i}(j) \sum_{u=1}^{d} E\left\{X_{t-k}^{2i} | s_{t-k} = u\right\} p_{uj}^{(k)} \pi(u)\,\sigma_{2i}\sigma_{2(l-i)}.\end{aligned}$$

By passing to vectorial representation the results follows. In particular,

$$\mu_{(4)}(k) = E\left\{X_t^4\right\} = \underline{\mathbf{1}}\left(I_{(d)} - \mathbb{P}^{(k)}\left(\sigma_4 a^4\right)\right)^{-1}\left(6\sigma_2^2 \mathbb{P}^{(k)}\left(a^2\right)\underline{\mu}_{(2)}(k) + \underline{\Pi}(\sigma_4)\right).$$

2. The proof of the second assertion is similar and is omitted.

□

Now, we carry out a second-order analysis of the squares ($Y_t = X_t^2$) of $MS-SBL$ which providing an alternate differentiation technique characterizing the processes $0-$dependent

Lemma 6. *For the second-order stationary $MS-SBL$, consider the process $Y_t = X_t^2$ and assume (to facilitate the exposition) that $k=2, m=1$ and for any $j \geq 0$, we set*

$$\underline{\mu}_Y(j) = \left(\pi(u) E\left\{Y_t Y_{t-j} | s_t = u\right\}, u = 1, \ldots d\right)',$$

then under the condition $\rho\left(\mathbb{P}^{(2)}\left(\sigma_2 a^2\right)\right) < 1$ we have

$$\underline{\mu}_Y(j) = \begin{cases} \underline{\mu}_{(4)}(2) & \text{if } j=0 \\ \left(I_{(d)} - \mathbb{P}\left(\sigma_2 a^2\right)\right)^{-1} \mathbb{P}(\sigma_2) \underline{\mu}_2(2) & \text{if } j=1 \\ \mathbb{P}^{(2)}\left(\sigma_2 a^2\right) \underline{\mu}_Y(n-2) + \mathbb{P}^{(n)}(\sigma_2) \underline{\mu}_2(2) & \text{if } j \geq n \geq 2 \end{cases},$$

and hence $\gamma_Y(j) = \mathbf{1}\underline{\mu}_Y(j) - \mu_2^2(2)$ for any $j \geq 0$.

Proof. The first part is obtained from the Lemma 5. For the second, $\pi(u) E\{Y_t Y_{t-1} | s_t = u,\} = a^2(u) \pi(u) E\{X_{t-2}^2 X_{t-1}^2$ $e_{t-1}^2 | s_t = u\} + \sigma_2 \pi(u) E\left\{X_{t-1}^2 | s_t = u\right\}$, so

$$\underline{\mu}_Y(1) = \mathbb{P}\left(\sigma_2 a^2\right) \underline{\mu}_Y(1) + \mathbb{P}\left(a^2\right) \underline{\mu}_{(2)}(2).$$

The third one is immediate. □

Remark 4. *According to Brockwell and Davies (1987) Section 3.2, it follows that if $(X_t)_t$ is a is $q-$dependent and $(e_t)_t$ is i.i.d. then for any measurable functions of $(X_t)_t$ is $q-$dependent. So, since because $\gamma_Y(j) \neq 0$ implies that the process $(Y_t)_t$ is not an $0-$dependent and hence $(X_t)_t$ is fare from an $MA(1)$ process.*

The next corollary we summarize some results listed above when $(s_t)_t$ is independent sequence noted hereafter $I-SBL$. The proof of this corollary is omitted as it consists of straightforward, but tedious algebra.

Corollary 1. *[Switching-independent case] Set $\bar{a} = E\{a(s_t)\}$ and $\lambda = \overline{a^2}\sigma_2$, then, for the $I-SBL$, we have:*

1. $\mu = 0$.
2. $\gamma_X(i) = \sigma_2(1-\lambda)^{-1} \delta_{\{i=0\}}$.
3. $\gamma_X(i,j) = \sigma_2^2\, \bar{a}(1-\lambda)^{-1} \delta_{\{i=k-m, j=k\}}$.
4. $E\{X_t^r\} = \left(\sigma_{2l} + \sum_{i=1}^{l-1} C_{2l}^{2i} \sigma_{2i} \sigma_{2(l-i)} \overline{a^{2i}} E\left\{X_t^{2i}\right\}\right)\left(1 - \sigma_{2l}\, \overline{a^{2l}}\right)^{-1} \delta_{\{r=2l\}}$.
5. $\gamma_Y(j) = \begin{cases} \dfrac{2\sigma_2}{\left(1-\sigma_2\overline{a^2}\right)\left(1-\sigma_4\overline{a^2}\right)}, & j=0, \\ \dfrac{2\sigma_2^3\overline{a^2}}{\left(1-\sigma_2\overline{a^2}\right)^2}, & j=1, \\ \sigma_2\overline{a^2}\gamma_Y(n-2), & j=n \geq 2. \end{cases}$

Remark 5. *The results of the fifth assertion of above corollary shows clearly that when $(s_t)_t$ is an independent sequence, the process $\left(X_t^2\right)_t$ associated to $k=2$ and $m=1$ has the same covariance structure of an $MS-ARMA(2,1)$, with an independent sequence $(s_t)_t$, i.e.,*

$$Z_t = \omega(s_t) + a_1(s_t)Z_{t-1} + a_2(s_t)Z_{t-2} + b(s_t)e_{t-1} + e_t \text{ with } a_1(.) = 0. \tag{12}$$

Indeed, the covariance structure of (12) is given by $\gamma_Z(j) = \mathbf{1}\underline{\mu}_Z(j) - \mu_Z^2$ where

$$\underline{\mu}_Z = \mathbf{1}\left(I_{(d)} - \mathbb{P}^{(2)}(a)\right)^{-1}\underline{\Pi}(\omega)$$

and

$$\underline{\mu}_Z(j) = \begin{cases} \left(I_{(d)} - \mathbb{P}^{(2)}\left(a_2^2\right)\right)^{-1}\left(\underline{\Pi}\left(\omega^2\right) + 2\mathbb{P}^{(2)}(\omega a_2)\underline{\mu}_Z + \underline{\Pi}(\sigma^2 b) + \underline{\Pi}(\sigma_2)\right), & j=0, \\ \left(I_{(d)} - \mathbb{P}(a)\right)^{-1}\left(\mathbb{P}(b)\underline{\Pi}(\sigma_2) + \mathbb{P}(\omega)\underline{\mu}_Z\right), & j=1, \\ \mathbb{P}(a_2)\underline{\mu}_Z(n-2) + \mathbb{P}^{(n)}(\omega)\underline{\mu}_Z, & j=n>1, \end{cases}$$

which is reduces in $I-SBL$ to $\gamma_Z(j) = \mu_Z(j) - \mu_Z^2$ where $\mu_Z = (1-\overline{a_2})^{-1}\ \overline{\omega}$ and

$$\gamma_Z(j) = \begin{cases} \sigma_2\left(1-\overline{a_2^2}\right)^{-1}(1+\overline{b^2}), & j=0, \\ \sigma_2\overline{b^2}(1-\overline{a_2})^{-1}, & j=1, \\ \overline{a_2}\gamma_Z(j-2), & j\geq 2. \end{cases}$$

Comparing $\gamma_Z(j)$ and that given in fifth assertion of corollary 1, it can be shown that for $\overline{\omega} = (1-\overline{a})\mu_{(2)}(k)$, $\overline{a_2} = \sigma_2\overline{a^2}$ and $\overline{b^2} = 1 + \sigma_4\overline{a^2}$, $\gamma_Z(j)$ and $\gamma_Y(j)$ coincide. This finding excludes the $I-SBL$ to being a weak white noise. ($0-$dependent). Moreover, the results of parts $1,2$ and 3 of the above Corollary are identical to the results given by Gabr (1988) when $d=1$. Furthermore, it is easy to see that the results in part 4 of Corollary 1 are in agreement with the results given by Martin (1999).

Example 2. Consider $d=2$, the entries of the transition matrix P are $p_{11}=\alpha$, $p_{22}=\beta$, the ergodic distribution is then $\pi(1) = (1-\alpha)/(2-\alpha-\beta)$. The kurtosis (in terms of $a(1)$ and $a(2)$) of $MS-SBL$ are shown in the Figure 1. The kurtosis of $MS-SBL$ in above figure are strictly positive and hence their distributions are leptokurtic i.e., the tails are thicker than normal one. Note here that all calculations thereafter have been performed with native code under $MATLAB$'15 computation language.

3.2 Diagonal model

Lemma 7. *(Second-order moments) If $(X_t)_t$ is generated by a $MS-DBL$, then under the condition $\rho\left(\mathbb{P}^{(k)}\left(\sigma_2 a^2\right)\right) < 1$ we have*

1. $\widetilde{\mu}_{(1)}(k) = E\{X_t e_t\} = \sigma_2$,
2. $\mu_{(1)}(k) = E\{X_t\} = \mathbf{1}\underline{\mu}_{(1)}(k)$ *where* $\underline{\mu}_{(1)}(k) = \mathbb{P}^{(k)}(a)\underline{\Pi}(\sigma_2)$,
3. $\widetilde{\mu}_{(2)}(k) = E\left\{X_t^2 e_t^2\right\} = \mathbf{1}\underline{\widetilde{\mu}}_{(2)}(k)$ *where* $\underline{\widetilde{\mu}}_2(k) = \left(I_{(d)} - \mathbb{P}^{(k)}\left(\sigma_2 a^2\right)\right)^{-1}\underline{\Pi}(\sigma_4)$ *and* $\mu_{(2)}(k) = E\left\{X_t^2\right\} = \mathbf{1}\underline{\mu}_{(2)}(k)$ *where* $\underline{\mu}_{(2)}(k) = \mathbb{P}^{(k)}\left(a^2\right)\underline{\widetilde{\mu}}_{(2)}(k) + \underline{\Pi}(\sigma_2)$.
4. $$\underline{\mu}_X(j) = \begin{cases} \mathbb{P}^{(k)}(a^2)\left(I_{(d)} - \mathbb{P}^{(k)}\left(\sigma_2 a^2\right)\right)^{-1}\underline{\Pi}(\sigma_4) + \underline{\Pi}(\sigma_2), & j=0, \\ \mathbb{P}^{(k)}(\ a)\mathbb{P}^{(k)}(\ 2\sigma_2)\underline{\Pi}(\sigma_2), & j=k, \\ \mathbb{P}^{(k)}(a)\mathbb{P}^{(n-k)}(\sigma_2)\underline{\mu}_{(1)}(k), & j=n>k. \end{cases}$$

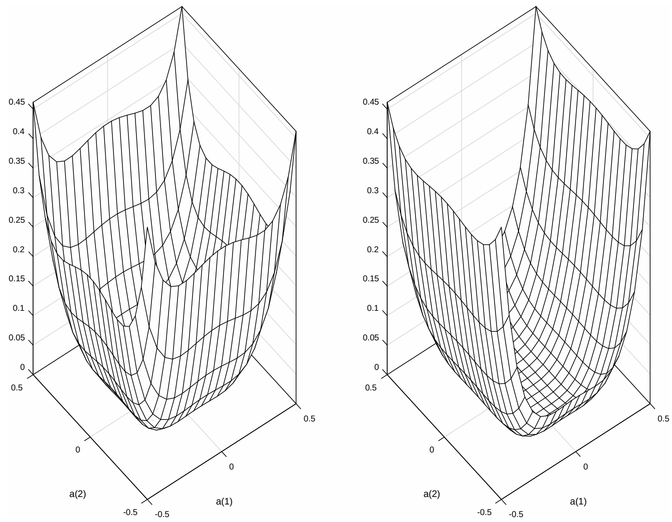

Figure 1: Kurtosis of the two-state $MS-SBL$. Left dependent $D-SBL$ and right $I-SBL$ with $\alpha = 0.75$ and $\beta = 0.95$.

Proof. The proof is straightforward and hence omitted. □

Note that in standard case $(d = 1)$ the process $(X_t)_t$ is an $k-$dependent and hence $\gamma(j) = 0$ for all $j \neq k$, in contrast here $\gamma(j) \neq 0$ for all j. This finding could be used as a sufficient condition for testing the Markovianity switching with dependent chain. Similar results stated in Lemma 5 for $MS-SBL$ may be given in $MS-DBL$ one.

Lemma 8. *For the $MS-DBL$ and under the condition $\rho\left(\mathbb{P}^{(k)}(\sigma_r a^r)\right) < 1$, we have*

1. $\widetilde{\mu}_{(r)}(k) = E\{(X_t e_t)^r\} = \mathbf{1}\underline{\widetilde{\mu}}_{(r)}(k)$ *where*

$$\underline{\widetilde{\mu}}_{(r)}(k) = \left(I_{(d)} - \mathbb{P}^{(k)}(\sigma_r\ a^r)\right)^{-1} (\sum_{i=1}^{\left[\frac{r}{2}\right]-1} C_r^{2i}\ \sigma_{2(r-i)}\mathbb{P}^{(k)}\left(\ \underline{a}^{2i}\right)\underline{\widetilde{\mu}}_{(2i)}(k) + \underline{\Pi}(\sigma_{2r})).$$

2. $\mu_{(r)}(k) = E\{X_t^r\} = \mathbf{1}\underline{\mu}_{(r)}(k) = \begin{cases} \underline{\mathbf{1}}\left(\underline{\Pi}(\sigma_{2n}) + \sum_{i=1}^{n} C_{2n}^{2i}\sigma_{2(n-i)}\mathbb{P}^{(k)}\left(\ \underline{a}^{2i}\right)\underline{\widetilde{\mu}}_{(2i)}(k)\right), & r = 2n, \\ \underline{\mathbf{1}}\sum_{i=1}^{n} C_{2n-1}^{2i-1}\ \mathbb{P}^{(k)}\left(\sigma_{2(n-i)}\ a^{2i-1}\right)\underline{\widetilde{\mu}}_{(2i-1)}(k), & r = 2n-1. \end{cases}$

Proof. For any integer $r \geq 1$, we have $(X_t e_t)^r = e_t^{2r} + \sum_{i=1}^{r} C_r^i a^i(s_t)(X_{t-k}e_{t-k})^i e_t^{2r-i}$, so

$$\begin{aligned}\pi(j)E\left\{(X_t e_t)^r | s_j = j\right\} &= E\left\{e_t^{2r}|s_t = j\right\}\pi(j) + \sum_{i=1}^{\left[\frac{r}{2}\right]} C_r^{2i} a^{2i}(j) E\left\{(X_{t-k}e_{t-k})^{2i}|s_t = j\right\}\pi(j)\sigma_{2(r-i)} \\ &= \sigma_{2r}\pi(j) + \sum_{i=1}^{\left[\frac{r}{2}\right]} C_r^{2i} a^{2i}(j) \sum_{u=1}^{d} E\left\{(X_{t-k}e_{t-k})^{2i}|s_{t-k} = u\right\} p_{uj}^{(k)}\pi(u)\sigma_{2(r-i)}\end{aligned}$$

and thus $\underline{\widetilde{\mu}}_{(r)}(k) = \left(I_{(d)} - \mathbb{P}^{(k)}(\ \sigma_r a^r)\right)^{-1}\left(\underline{\Pi}(\sigma_{2r}) + \sum_{i=1}^{\left[\frac{r}{2}\right]-1} C_r^{2i}\ \sigma_{2(r-i)}\mathbb{P}^{(k)}\left(\ \underline{a}^{2i}\right)\underline{\widetilde{\mu}}_{(2i)}(k)\right)$. Therefore

$$E\{(X_t e_t)^r\} = \mathbf{1}\underline{\widetilde{\mu}}_{(r)}(k).$$

For the second assertion, we have $X_t^r = e_t^r + \sum_{i=1}^{r} C_r^i a^i(s_t)(X_{t-k}e_{t-k})^i e_t^{r-i}$, so if $r = 2n$, we get $\pi(l) E\left\{X_t^{2n} | s_t = l\right\}$ $= \sigma_{2n}\pi(l) + \sum_{i=1}^{n} C_{2n}^{2i}\sigma_{2(n-i)} a^{2i}(l) E\left\{(X_{t-k}e_{t-k})^{2i} | s_t = l\right\}\pi(l)$. Hence

$$\underline{\mu}_{(2n)}(k) = \underline{\Pi}(\sigma_{2n}) + \sum_{i=1}^{n} C_{2n}^{2i}\ \sigma_{2(r-i)}\mathbb{P}^{(k)}\left(\ a^{2i}\right)\underline{\widetilde{\mu}}_{(2i)}(k)$$

otherwise,

$$\underline{\mu}_{(2n-1)}(k) = \sum_{i=0}^{n-1} C_{2n-1}^{2i-1}\ \sigma_{2(r-i)}\mathbb{P}^{(k)}\left(\ a^{2i-1}\right)\underline{\widetilde{\mu}}_{(2i-1)}(k),$$

so $E\{X_t^r\} = \mathbf{1}\underline{\mu}_{(r)}(k) I_{\{r=2n\}} + \mathbf{1}\underline{\mu}_{(r)}(k) I_{\{r=2n-1\}}$. □

All the third-order moments of the $MS-DBL$ are summarized in the following lemma

Lemma 9. *(Third-order moments) If $(X_t)_t$ is generated by a $MS-DBL$, then under the condition of Lemma 8, the non centred third-order moments are*

$$\underline{\mu}_X(i,j) = \begin{cases} 3\sigma_2\underline{\mu}_{(1)}(k) + \ \mathbb{P}\left(a^3\right)\underline{\widetilde{\mu}}_{(3)}(k), & i=j=0, \\ \mathbb{P}^{(k)}\left(a^2\right)\left\{3\sigma_4\underline{\mu}_{(1)}(k) + \mathbb{P}^{(k)}\left(\sigma_2 a^3\right)\underline{\widetilde{\mu}}_{(3)}(k)\right\} + \mathbb{P}^{(k)}(\sigma_2)\underline{\mu}_{(1)}(k), & i=0, j=k, \\ \mathbb{P}\left(a^2\right)\left\{\mathbb{P}^{(k(n-1))}\left(2\sigma_2^2\right)\underline{\mu}_{(1)}(k) + \sigma_2\underline{\mu}_X(0,k(n-1))\right\} & \\ \qquad\qquad + \mathbb{P}^{(nk)}(\sigma_2)\underline{\mu}_{(1)}(k), & i=0, j=nk, n\geq 2 \\ \mathbb{P}^{(k)}(a)\left\{\mathbb{P}^{(k)}\left(3\sigma_2 a^2\right)\widetilde{\underline{\mu}}_{(2)}(k) + \underline{\Pi}(\sigma_4)\right\}, & i=j=k, \\ 4\sigma_2^2\mathbb{P}^{2(k)}(a)\underline{\mu}_{(1)}(k), & i=k, j=2k, \\ 2\sigma_2\mathbb{P}^{2(k)}(a)\mathbb{P}^{(k(n-2))}(\sigma_2)\underline{\mu}_{(1)}(k), & i=k, j=nk, n\geq 3 \\ \mathbb{P}^{(k)}(a)\mathbb{P}^{(k(l-1))}(\sigma_2)\underline{\mu}_X(k(n-l))), & i=lk, j=nk, l\geq 2, n\geq 3, \end{cases}$$

and hence $\gamma_Y(i,j)$ may be expressed from (11).

Proof. The first part corresponding to $i=j=0$ may be derived through Lemmas 7 and 8. For the case: $i=j=k>0$, then we have $\pi(u) E\left\{X_t X_{t-k}^2 | s_t = u\right\} = \pi(u) a(u) E\left\{X_{t-k}^3 e_{t-k} | s_t = u\right\}$, so $\underline{\mu}_X(k,k) = \mathbb{P}^{(k)}(a)\underline{U}(k)$ where the components of the vector $\underline{U}(k)$ are $\pi(u) E\left\{X_t^3 e_t | s_t = u\right\} = \pi(u)\sigma_4 + 3\pi(k)\sigma_2 a^2(u) E\left\{X_{t-k}^2 e_{t-k}^2 | s_t = u\right\}$ and we deduce that $\underline{U}(k) = \underline{\Pi}(\sigma_4) + \mathbb{P}^{(k)}\left(3\sigma_2 a^2\right)\widetilde{\underline{\mu}}_{(2)}(k)$ hence the results follows. Similar methodology can be used to obtain the expansion of the remainder cases. □

Lemma 10. *For the $MS-DBL$ consider the squared process $Y_t = X_t^2$ and assume that $\rho\left(\mathbb{P}^{(k)}(\sigma_4 a4)\right) < 1$, then we have*

$$\underline{\mu}_Y(j) = \begin{cases} \underline{\mu}_{(4)}(k), & j=0, \\ \mathbb{P}^{(k)}(a^2)\left\{\underline{\Pi}(\sigma_6) + 6\sigma_4\mathbb{P}^{(k)}\left(a^2\right)\underline{\widetilde{\mu}}_{(2)}(k) + \sigma_2\mathbb{P}^{(k)}\left(a^4\right)\underline{\widetilde{\mu}}_{(4)}(k)\right\} + \mathbb{P}^{(k)}(\sigma_2)\underline{\mu}_{(2)}(k), & j=k, \\ \mathbb{P}^{(k)}(\sigma_2 a^2)\underline{\mu}_Y(k(n-1)) + \left(\mathbb{P}^{(k)}(a^2)\mathbb{P}^{(k(n-1))}\left(2\sigma_2^2\right) + \mathbb{P}^{(kn)}(\sigma_2)\right)\underline{\mu}_{(2)}(k), & j=nk, n\geq 2. \end{cases}$$

and hence $\gamma_Y(j) = \mathbf{1}\underline{\mu}_Y(j) - \mu_{(2)}^2$ for any $j\geq 0$.

Proof. The first part, for $j=0$ may be deduced from the Lemma 8. The second part, for $j=k$, let us define the vector $\underline{\mu}_Y(k) = \left(\pi(j) E\left\{X_t^2 X_{t-k}^2 | s_t = j\right\}, 1\leq j\leq d\right)'$, then its is not difficult to see that

$\underline{\mu}_Y(k) = \mathbb{P}^{(k)}\left(a^2\right)\underline{W}(k) + \mathbb{P}^{(k)}(\sigma_2)\underline{\mu}_{(2)}(k)$ where $\underline{W}(k) = \left(\pi(j)E\left\{X_t^4 e_t^2 | s_t = j\right\}, 1 \leq j \leq d\right)'$. Tedious calculation shows that $\underline{W}(k) = \underline{\Pi}(\sigma_6) + \mathbb{P}^{(k)}(6\sigma_4 a^2)\underline{\widetilde{\mu}}_{(2)}(k) + \mathbb{P}^{(k)}(\sigma_2 a^4)\underline{\widetilde{\mu}}_{(4)}(k)$ and hence the results follow. The third part, follow upon the observation that $\underline{\mu}_Y(kn) = \mathbb{P}^{(k)}(a^2)\underline{U} + \mathbb{P}^{(nk)}(\sigma_2\underline{\mu}_{(2)}(k)$, where

$$\underline{U} = (\pi(j)E\left\{X_t^2 X_{t-k(n-1)}^2 | s_t = j\right\}, 1 \leq j \leq d)',$$

after some computations, we get $\underline{U} = \sigma_2\underline{\mu}_Y(k(n-1)) + \mathbb{P}^{(k(n-1))}(2\sigma_2)\underline{\mu}_{(2)}(k)$ and hence the result follows. □

The following corollary correspond to independent switching $MS-DBL$, we stated the moments properties without proofs. Details of the proof are of course available if it required.

Corollary 2. [*Switching-independent case*] *For the $I-DBL$, we set $\lambda = \sigma_2\overline{a^2}$, then we have:*

1. $\mu = E\{X_t\} = \sigma_2\overline{a}$ *and* $Var(X) = \sigma_2 + \dfrac{\overline{a^2}\sigma_4}{1-\lambda} - \mu^2$.
2. $\gamma_X(i) = \lambda^2\ \delta_{\{i=k\}}$.
3. $$\gamma_X(i,j) = \begin{cases} \frac{3\overline{a^3}\sigma_4}{(1-\lambda)}\left(\overline{a^2}\sigma_4 - \sigma_2\right) + \overline{a^3}\left(\ \sigma_6 - 2\sigma_2^3\right), & i = j = 0, \\ \frac{\overline{a^3}\sigma_2}{\left(1-\sigma_2\overline{a^2}\right)} - \beta - 2\overline{a^3}\sigma_2^3, & i = 0, j = k \\ \frac{\overline{a^{2n+1}}\sigma_2^{2n}}{\left(1-\sigma_2\overline{a^2}\right)}\beta, & i = 0, j = nk, n = 2, 3, ... \\ \frac{\overline{a}}{\left(1-\sigma_2\overline{a^2}\right)}\left(\ \sigma_4 - \sigma_2^2 + \overline{a^2}\ \left(\sigma_4 - \sigma_2^2 + 2\overline{a^2}\sigma_2^3\right)\right), & i = j = k,,. \\ \overline{a^3}\sigma_2^3, & i = k, j = 2k, \\ 0, & otherwise \end{cases}$$
   where $\beta = \overline{a^4}\left(3\sigma_4^2 - \sigma_2\sigma_6\right) + \overline{a^2}(\sigma_6 + \sigma_2\sigma_4) + 2\sigma_4$.
4. $$\gamma_Y(j) = \begin{cases} E\left\{X_t^4\right\} - \ \mu_{(2)}^2, & j = 0 \\ \overline{a^2}\left\{\sigma_6 + 6\sigma_4\overline{a^2}\widetilde{\mu_{(2)}} + \sigma_2\overline{a^4}\widetilde{\mu}_{(4)}\right\} + \sigma_2\mu_{(2)} - \mu_{(2)}^2, & j = k, \\ \sigma_2\overline{a^2}\gamma_Y(k(n-1)) + \left(2\sigma_2\overline{a^2} + \sigma_2^2\right)\mu_{(2)} + \sigma_2\overline{a^2}\mu_{(2)}^2\ \text{-}\mu_{(2)}^2\ \ if\ j = nk, & n \geq 2. \end{cases}$$

Remark 6. *The results of second assertion of above corollary shows clearly that $I-DBL$ is a $k-$dependent process and hence it has the same covariance structure of an $MS-MA(k)$ model, in other hand the result of the fourth assertion clearly indicate that $I\ -DBL$ is not an $MS-MA(k)$.*

Remark 7. *The results obtained for $MS-BL$ can be useful in modeling nonlinear time series, particularly in the choice of the lags k and l of some simple bilinear models for which the innovation is Gaussian. Note here that the results obtained in independent switching case are similar to standard ones $(d = 1)$ when the parameters in moment properties are replaced by the moments of their coefficients. Despite the difficulty and the complexity of computation, especially in dependent switching, the results obtained here constitute an alternative for the study of nonlinear series.*

Example 3. Consider the same distribution as in Example 2, then the kurtosis and the skewness of $MS-DBL$ are shown in the Figure 2 terms of $a(1)$ and $a(2)$) of $MS-SBL$ are shown in the Figure 1. The kurtosis of $MS-DBL$ in top panels of the Figure 2 are strictly positive and hence their distributions are leptokurtic. The skewness of $MS-DBL$ in bottom panels are negative for some values of $a(1)$ and $a(2)$ and hence the distribution has a long tail to the left, i.e., mode>median>mean.

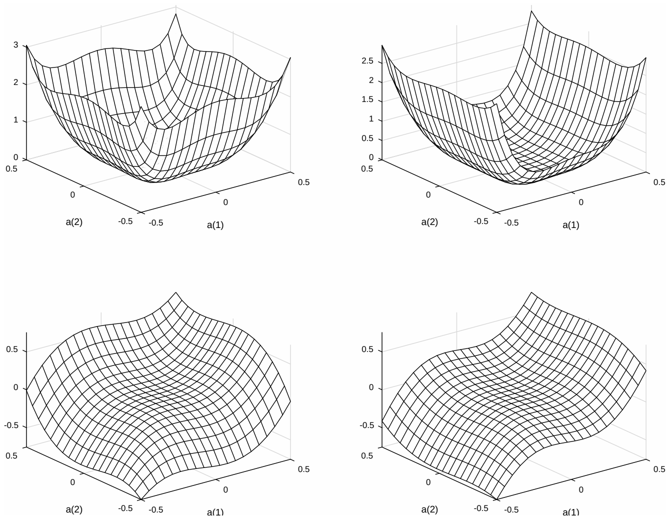

Figure 2: Kurtosis and skewness of the two state $MS-DBL$. The left of top panel displayed the kurtosis of $D-DBL$ followed by $I-DBL$. Bottom panel displayed the skewness of $D-DBL$ in left followed by $I-DBL$.

3.3 Simulation study

In order to examine the problems associated with the identification of a $MS-BL$ model with the help of the third or fourth order moment, we consider an $MS-SBL$ (resp $MS-DBL$) with $(e_t)_t$ is a Gaussian sequence $\mathcal{N}(0,1)$ and $(s_t)_t$ is a 2−state Markov chain with $p_{11}=0.75$, $p_{22}=0.95$ and $a(s_t)\in\{-0.75,0.15\}$. Now we simulate a series of length $n=1000$, according to the models already mentioned, and calculate the second, third and fourth order moments $\widehat{\gamma}_X(i)$, $\widehat{\gamma}_X(i,j)$, and $\widehat{\gamma}_{X^2}(i)$, $i,j=1,...,\max lag$. We repeated the experiment 500 times and the mean values of $\widehat{\gamma}_X(i)$, $\widehat{\gamma}_X(i,j)$ and $\widehat{\gamma}_{X^2}(i)$ are obtained. The simulation experiment has been made in order to compare the difference between the theoretical values of $\gamma_X(i)$, $\gamma_X(i,j)$ and $\gamma_{X^2}(i)$ and those obtained via a Monte Carlo experiment. To avoid biased simulation's intervention, we delete first 100 observations to make sure the randomness of the model. The different plots of models according to the variability of the chain $(s_t)_t$ are presented in Figure 3. In Figure 3, we can see that the graphic associated to $MS-SBL$, presents more volatility than the $MS-DBL$. Moreover, it can also be observed that there is a closed similarity between the $C-.$ and $I-.$ models.

3.4 Second-order properties

The sample $\widehat{\gamma}_X(s)$ up to lag 25, according to the variability of the chain $(s_t)_t$ are given in Figure 4. As expected theoretically, the covariances in all cases decay to zero as the lag increases, though the rate of decay in the case of the $MS-DBL$ is slower. The results of the simulation of the second-order moment are given in Table 1.

Observations of the series simulated from $MS-SBL$ (resp. $MS-DBL$) are plotted in right (resp. left) side panel of Figure 3 for different specification of the chain $(s_t)_t$. The covariance $\gamma_X(h)$ of these series are shown in Table 1, for $h\in\{0,1,2,3,4,5\}$ and their plots are in Figure 4. So, we can see that the covariance function are zero beyond 0 (resp. 1) for $MS-SBL$(resp. $MS-DBL$). This finding confirm the theoretical results summarized in Figure 4.

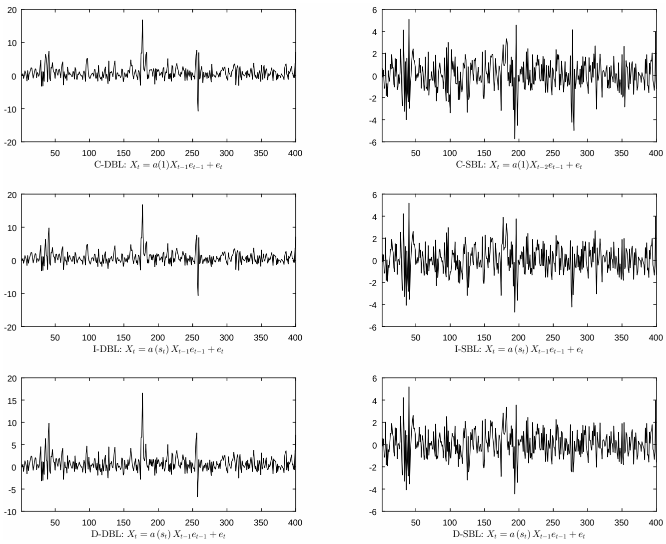

Figure 3: Displayed are in top panel $C-DBL$ followed by $C-SBL$ in second panel the $I-DBL$ followed by $I-SBL$ and in the third $D-DBL$ followed by $D-SBL$.

3.5 Third-order properties

Now, the third−order moments are concerned. The sample third-order moments $\widehat{\gamma}_X(0,s)$ up to lag 25, according to the variability of the chain $(s_t)_t$ are given in Figure 5. It can be observed that for the $MS-SBL$, $\widehat{\gamma}_X(0,s)$ are approximately zero, contrary to the $MS-DBL$, $\widehat{\gamma}_X(0,s)$ are not zero even beyond 1. This means that the white noise structure of $MS-SBL$ is preserved by the third-order moment properties and the structure $MA(1)$ of $MS-DBL$ is violated. The results of simulation of the third-order moments are shown in Table 2 for $MS-SBL$ when $(s_t)_t$ is an independent (resp. dependent) chain, and in Table 2 for $MS-DBL$ with the same structure chain followed by their true values (results between parentheses). Table 2 gives theoretical values (in the parenthesis) of $\gamma(i,j)$ for different values of i,j and the corresponding simulated values $\widehat{\gamma}(i,j)$ for $MS-DBL$.

Observations of the first series simulated from $MS-SBL$ is plotted in first panel of Figure 3. Its second-order $\gamma_X(m)$ and third-order $\gamma_X(i,j)$ moments are shown respectively in first column of Table 1 and in Table 2 with $(i,j) \in \{0,1,2,3,4\} \times \{0,1,2,3,4\}$. It can be seen that simulation results agree well with the results reported in in Figure 5 and with the theoretical results. Moreover, the order of $\gamma_X(i,j)$ and $\gamma_X(m)$ are quite different, we believe that the simple bilinear structure are not obvious and hence, an AIC criterion should be discussed further.

Observations of this series (when $(s_t)_t$ is independent) simulated from $MS-SBL$ is plotted in second panel of Figure 1. Its second-order $\widehat{\gamma}_X(m)$ and third-order $\widehat{\gamma}_X(i,j)$ moments are shown respectively in second block of Table 1 and in Table 2. In this case also the simulation results are quite closely with the theoretical results. Moreover, the values corresponding to the cell $(0,0)$ are found to be absolutely dominant over other values. Since $\gamma_X(2)$ is zero, we believe that $\gamma_X(2,2)$ should be wrong message.

3.6 Covariance analysis of squared $MS-SBL$ and $MS-DBL$

The purpose of this subsection is to carry out a second-order analysis on the squared $MS-SBL$ and $MS-DBL$ noted $Y_t = X_t^2$ with the help of the Lemmas 6 and 10 to providing an alternate differentiation technique from their linear representations

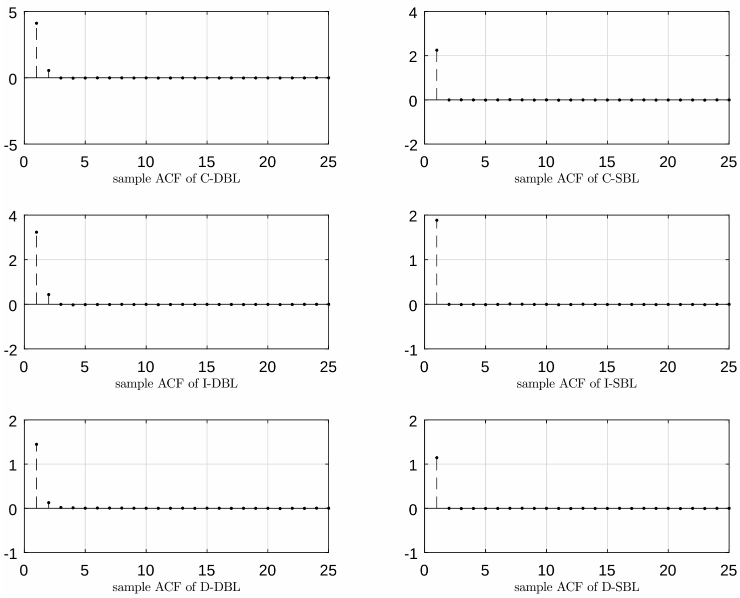

Figure 4: From top to bottom: The covariance functions associated respectively to $(.)-DBL$ followed in right side by $(.)-SBL$

Observations of this series simulated from the squared version of $MS-SBL$ and $MS-DBL$ their covariance $\widehat{\gamma}_Y(m)$ is plotted in Figure 6. and there values are shown in Table 4.

As a consequence of the results reported in Table 4 is that the second-order structure of white-noise for $MS-SBL$ model (resp. $MA(1)$ for $MS-DBL$) are not preserved by the process $(Y_t)_t$. So, the squared process maybe constituted a powerful alternative for the differentiation between linear and/or nonlinear models. In the other words, in view of Figure 6, it can be see that the monotony of $\widehat{\gamma}_X(h)$ is also preserved by $\widehat{\gamma}_{X^2}(h)$ for $h = 0, 1, \ldots$. Note here that in spite the large values of $\gamma_{X^2}(h)$ for $C-DBL$ and $I-DBL$, the values of $\gamma_{X^2}(h)$ corresponding to $D-DBL$ are significantly relaxed.

4 Application to investments

In this section, we use our theoretical results to analyze certain series of asset returns via $MS-BL$ models (9) aimed to predict the future values of such series. The resort to regime switching in asset returns and their pronounced implications for investments have been widely documented through several models according some accommodates features. (see Marcucci (2005) for more discussions). Based on the findings in this literature, we analyze the exchange rates of the Algerian Dinar against the single European currency (Euro) (EUR/DZD). We investigate some descriptive statistics and the impact of such series for the different regimes. Our base model is an $MS-BL$ models (9) with two regimes are considered. We compare its implications to those of restricted models to determine the impact of regime switching models.

4.1 Sample data and preliminary analysis

The proposed models $MS-SBL$ and $MS-DBL$ are investigated to model the series of EUR/DZD already studied by Bibi (2021), observed from January 3, 2000 to September 29, 2011. Since there is some weeks

Table 1: The Monte Carlo simulations and theoretical values (in parenthesis) of $\gamma_X(s)$ in different lags of MS-SBL and MS-DBL model.

Lags\Models	$MS-SBL$			$MS-DBL$		
	$C-SBL$	$I-SBL$	$D-SBL$	$C-DBL$	$I-DBL$	$D-DBL$
0	2.2878 (2.2857)	1.8946 (1.8957)	1.4419 (1.4269)	4.0592 (4.0946)	3.2296 (3.2639)	1.4608 (1.4594)
1	−0.0042 (0.0000)	−0.0022 (0.0000)	−0.0006 (0.0000)	0.5219 (0.5191)	0.4200 (0.4192)	0.1216 (0.0984)
2	0.0017 (0.0000)	−0.0068 (0.0000)	−0.0059 (0.0000)	−0.0400 (0.0000)	−0.0249 (0.0000)	0.0053 (0.0045)
3	−0.0031 (0.0000)	−0.0038 (0.0000)	−0.0038 (0.0000)	−0.0583 (0.0000)	−0.0498 (0.0000)	−0.0010 (0.0006)
4	−0.0095 (0.0000)	−0.0082 (0.0000)	−0.0042 (0.0000)	−0.0453 (0.0000)	−0.0378. (0.0000)	−0.0091 (0.0080)
5	−0.0026 (0.0000)	−0.0031 (0.0000)	−0.0036 (0.0000)	−0.0305 (0.0000)	−0.0315 (0.0000)	−0.0057 (0.0045)

comprise less than five observations (due to legal holidays), we remove the entire week with less than five available. So, the final length of the series is $T = 3055$ observations uniformly disturbed on 611 weeks. The elementary statistics are tabulated in Table ($Stat$) below which shows some description of the series EUR/DZD. The kurtosis is less than the normal value of 3 indicating that the distribution has lighter tails and a flatter peak than a normal distribution (platykurtic). The skewness is significantly negative, showing that the tail on the left side of the histogram is longer than the right, with most of the data clustered on the right side of the mean.

Note: The $Jbstat$-test which has a χ^2 distribution with 2 degrees of freedom under the null hypothesis of normally distributed errors. The 5% critical value is therefore 6.4795. The $LM(12)$ statistic is the $ARCHLM$ test up to the twelfth lag and under the null hypothesis of no $ARCH$ effects it has a $\chi^2(q)$ distribution, where q is the number of lags. The $Q^2(12)$ statistic is the Ljung-Box test on the squared residuals of the conditional mean regression up to the twelfth order. Under the null hypothesis of no serial correlation, the test is also distributed as a $\chi^2(q)$ where q is the number of lags. Thus, for both tests the 5% critical value are 6.5597 (resp.31.4104). At a confidence level of 5% both skewness and kurtosis are significant, since the standard errors under the null of normality are $6/T = 0.0019$ and $24/T = 0.0077$ respectively.

The plot of the trajectory of EUR/DZD series followed by its second and third-order cumulants are given in the following Figure 7

It is clear that the series displayed in Figure 7 exhibit a structural break changes. More precisely, from the results of Table 6 we can see that the descriptive statistics show that the negativity skewed, indicate that the most data points are clustered on the right side (higher values), with a few extreme low values stretching out to the left. So, we have mean < median < mode. Moreover, the positivity of the kurtosis shows that the data are Leptokurtic and hence the null hypothesis of normality is rejected. Additionally, the Jarque-Bera test statistics ($Jbstat$) is larger than the critical value (6.5597) which is significant of 1% level which confirm the rejection of null hypothesis. The plot of third-order cumulant displayed in Figure 7 performs the rejection.

Table 2: The third-order moments of I-SBL and D-SBL models.

$\frac{Models}{Lags}$	$I-SBL$ model					$10^{-3}\times D-SBL$ models				
	0	1	2	3	4	0	1	2	3	4
0	-0.0007 (-0.0006)	0.0002 (0.0000)	-0.0009 (-0.0007)	0.0003 (0.0005)	0.0002 (0.0000)	-0.0363 (-0.0291)	0.1553 (0.1522)	-0.0265 (-0.0255)	0.0154 (0.0164)	-0.1085 (-0.1001)
1		0.0003 (0.0000)	0.0019 (0.0020)	0.0001 (0.0000)	-0.0001 (0.0000)		0.0820 (0.0789)	0.5082 (0.4811)	-0.1040 (-0.1003)	-0.0599 (-0.0411)
2			-0.0007 (-0.0006)	0.0001 (0.0000)	0.0003 (0.0001)			0.0641 (0.0788)	0.0369 (0.0335)	-0.0614 (0.0059)
3				-0.0002 (-0.0005)	-0.0002 (0.0000)				-0.0310 (0.0325)	0.0295 (0.0198)
4					-0.0003 (0.0000)					-0.0252 (-0.0250)

Table 3: The third-order moments of I-DBL and D-DBL models

$\frac{Models}{Lags}$	$I-DBL$ model					$D-DBL$ model				
	0	1	2	3	4	0	1	2	3	4
0	−8.1885 (−8.09871)	3.7328 (3.7294)	−0.4178 (−0.3987)	−0.3624 (−0.3592)	−0.2654 (−25904)	1.9461 (1.9392)	1.0668 (1.0590)	−0.0452 (−0.0450)	−0.0528 (−0.0492)	0.0540 (0.0601)
1		5.9508 (5.8972)	0.2878 (0.2798)	−0.1515 (−0.1405)	−0.2007 (−0.2001)		1.2645 (1.2551)	0.0826 (0.0802)	0.0283 (0.0279)	−0.1077 (−0.1009)
2			3.0620 (3.1001)	−0.0706 (−0.0710)	−0.0334 (0.0335)			0.6148 (0.5941)	0.0414 (0.0092)	−0.0323 (0.0325)
3				1.2265 (1.2155)	−0.0571 (0.0493)				0.2780 (0.2689)	−0.0415 (−0.0491)
4					0.5082 (0.4905)					0.1569 (0.1601)

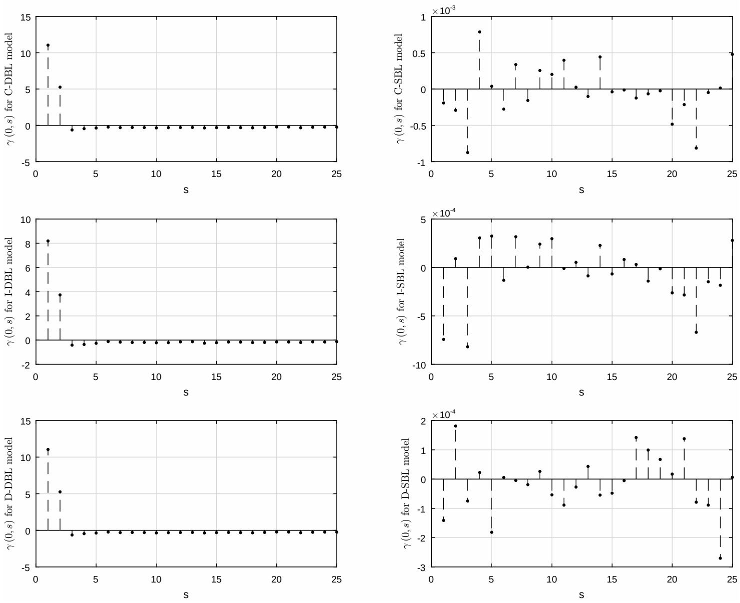

Figure 5: Displayed are the sample third-order moments $\widehat{\gamma}_X(0,s)$ of the $MS-DBL$ in left panel and of $MS-SBL$ in right.

Table 4: The simulate values of γ_{X^2} for different values of lags for the MS-SBL and MS-DBL.

Models / *Lags*	*Squared* $MS-SBL$			*Squared* $MS-DBL$		
	$C-DBL$	$I-DBL$	$D-DBL$	$C-DBL$	$I-DBL$	$D-DBL$
0	46.2211	19.4792	4.3941	405.3164	206.3111	40.9966
1	5.9235	3.4578	0.6353	222.7964	95.6614	18.5253
2	23.4193	8.2800	1.0314	113.4164	38.2723	7.6013
3	3.6595	1.7531	0.2424	56.5737	15.7618	4.0707
4	11.4029	3.4330	0.3205	32.3876	7.0561.	3.8038
5	1.8171	0.7900	0.0889	15.4928	3.3257	3.2314

4.2 Fit $MS-BL$ to daily EUR/DZD

Based on the above description of EUR/DZD series, the $MS-ARMA$ model did not reflect the behavior of such data, and hence $MS-BL$ models may be called for modelling and forecasting this series. Indeed, for modelling purpose, we are firstly using a quasi-maximum likelihood (QML) procedure for fitting the series by $MS-BL$. Table 6 some goodness-of-fit statistics are reported according to C_BL. These statistics are used as model selection criteria. The Akaike information criterion (AIC) and the Schwarz criterion (BIC) (results not reported here) both indicate that the best model is the $MS-DBL$. Another property of $MS-DBL$ models that emerges from Table 6 is the high persistence showed by small parameters estimates. The results of the parameter estimation followed by the mean squares errors (results between parentheses), according to $C-SBL$ and $C-DBL$ models are reported in the column (Parameters) of Table 6. The results gathered in (statistics column) of Table 6 are the elementary statistics associated with the adjusted series

This table shows the parameter estimates of adjusted EUR/DZD by $MS-BL$ under $d=1$ and their

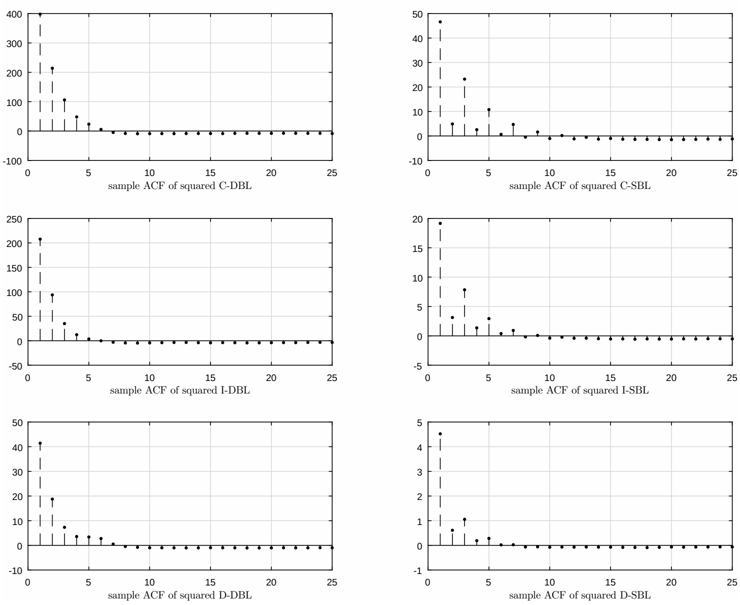

Figure 6: Displayed are the $\widehat{\gamma}_Y(s)$ of the squared $MS-DBL$ in left panel and of squared $MS-SBL$ in right one.

Table 5: Elementary statistics of daily EUR/DZD series.

Statistics:	*Mean*	*Std*	*Skew*	*Kur*	*Jbstat*	$LM(12)$	$Q^2(12)$
10^4∗Results	0.0088	0.0011	-0.0001	0.0002	0.0232	0.2688	5.8687

Table 6: The elementary statistics of adjusted $MS-BL$ under $d=1$.

Models	Parameters	10^4∗*Statistics*						
		Mean	*Std*	*skew*	*kur*	*Jbstat*	$LM(12)$	$Q^2(12)$
$C-SBL$	0.099 (0.045)	0.0088	0.0011	-0.0008	0.0004	0.0548	0.2720	0.8480
$C-DBL$	0.096 (0.031)	0.0089	0.0012	-0.0001	0.0003	0.0211	0.2697	1.6060

elementary statistics. The main observation is that the values of elementary statistics of the series adjusted by $C-DBL$ model are better than $C-SBL$ comparing with the original series. This is not surprising observation because the model $C-SBL$ is a weak white noise whereas the EUR/DZD series is very far to be considered as a white noise. The second approach for modelling the EUR/DZD is by $MS-$ BL under $d=2$ with independent and dependent chain, are summarized in Table 7.

In above Table 7 the column "Matrix P" means the estimate of transition matrix of the chain. Now, a few comments can be made. Indeed, it is clear that the Table 7 demonstrates that the $MS-DBL$ model applied to the EUR/DZD series performs as effectively as the applications to EUR/DZD concerning classification certainty for the mean. At least 96% of the fitted observations are considered with low

Table 7: The elementary statistics of adjusted $MS-BL$ under $d=2$.

Models	Parameters	$10^4 * Statistics$							Matrix P
		$Mean$	Std	$skew$	kur	$Jbstat$	$LM(12)$	$Q^2(12)$	
$I-SBL$	0.0103 0.0097 (0.021 0.014)	0.0088	0.0011	−0.0008	0.0004	0.0550	0.2720	0.8480	$\begin{pmatrix} 0.5420 & 0.4580 \\ 0.5416 & 0.4584 \end{pmatrix}$
$D-SBL$	0.0080 0.0117 (0.037 0.024)	0.0088	0.0011	−0.0009	0.0004	0.0521	0.2718	0.8573	$\begin{pmatrix} 0.3132 & 0.6868 \\ 0.3021 & 0.6979 \end{pmatrix}$
$I-DBL$	0.0115 0.0086 (0.011 0.015)	0.0089	0.0012	−0.0001	0.0003	0.0211	0.2697	1.6060	$\begin{pmatrix} 0.9979 & 0.0021 \\ 0.9979 & 0.0021 \end{pmatrix}$
$D-DBL$	0.0090 0.0121 (0.087 0.064)	0.0088	0.0019	−0.0008	0.0004	0.0521	0.2718	0.8573	$\begin{pmatrix} 0.3132 & 0.6868 \\ 0.6320 & 0.3680 \end{pmatrix}$

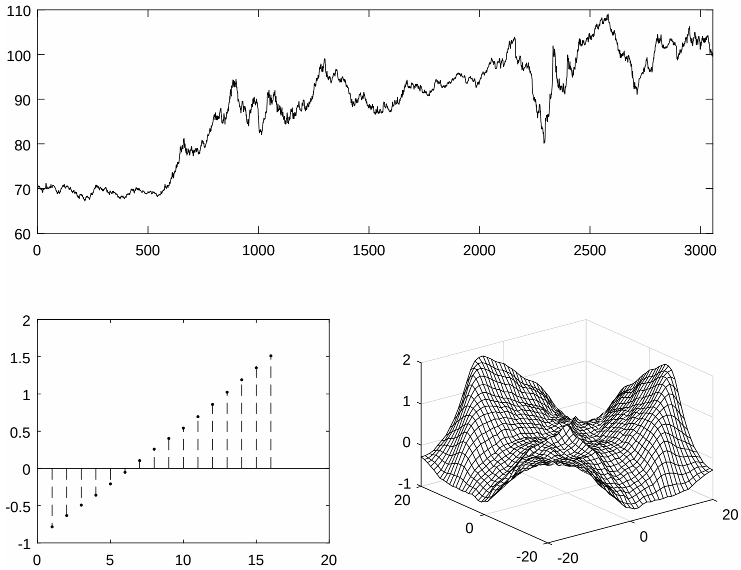

Figure 7: Top panel: the EUR/DZD series. Bottom panels: the second (left) and third (right) order cumulant of EUR/DZD series.

uncertainty. In terms of variance, the model classify at most 95% of the observations as decisive. The $MS-DBL$ with a greater number of states, applied to EUR/DZD is the least effective in classifying variance, with a percentage of "decisive" between 99% and 100%. Regarding the skewness and the kurtosis are the same as for original data with non-significant difference The $Jbstat$-test is also significantly to reject the null. In end, the $ARCH$ effect tests indicate clearly that no serial correlation up to the twelfth lag. Nonetheless, the model is indicated as good alternatives in terms of goodness− of− fit. According to the study, the $MS-DBL$ with $d>2$, is a reasonable choice for the application. This conclusion is obtained when accounting for both "classification power" and goodness-of -fit.

5 Conclusion

In this paper, we propose a unified framework that enables the calculation of higher-order moments and the diagnostic analysis of an $MS\text{-}BL(p,0,P,Q)$ model under both independent and dependent switching structures. As a crucial first step, we represent the model in a Markovian state-space form, which allows us to derive the conditions for the existence of strict stationary and ergodic solutions in $\mathbb{L}_p$, $p\geq 1$, under which explicit expressions for the first-, second-, and third-order moments are obtained. The second step of our framework focuses on analyzing specific cases of $MS\text{-}BL(p,0,P,Q)$ models. More precisely, the first-order $MS\text{-}SBL$ and $MS\text{-}DBL$ models are examined in detail.

For the $MS\text{-}SBL$ model, the first- and second-order moments are derived under both (in-)dependent switching schemes, which exhibit structural similarities to those of an $MS\text{-}MA(1)$ model. Consequently, studying the second-order structure of the squared process allows us to distinguish between $MS\text{-}SBL$ and $MS\text{-}MA(1)$ models. The same methodology is applied to the $MS\text{-}DBL$ model. Our main theoretical contribution lies in demonstrating how this generalized formulation can be used to derive the first and second moments conditional on dependent regime switching. In contrast, for the $MS\text{-}DBL$ model with independent regime switching, the first- and second-order moments coincide with those of the standard model and can be identified under covariance with the $MS\text{-}MA(1)$ model. However, this equivalence does not extend to the third-order moment structure, which breaks this covariance-based similarity.

Nonetheless, examining the covariance structure of squared processes can help resolve this ambiguity.

In this study, we have made the following assumptions:
(i) $d = 2$, since the implications of the transition matrix prevent us from considering higher-order models, even in cases such as $d = 3$ and $m = 4$ (in MS-SBL).
(ii) A lower-triangular model is adopted to obtain a Markovian representation that simplifies the analytical derivations.
(iii) The restriction to a model without a moving average component is justified by the fact that the stationarity and ergodicity conditions are independent of this part.
(iv) Gaussian innovations are assumed to simplify otherwise cumbersome computations.
(v) The estimation of cumulants is known to be biased.

These specifications introduce some limitations to our analysis. Nonetheless, the empirical results demonstrate that ***MS-DBL*** models significantly outperform standard ***BL*** models according to a broad range of elementary statistical criteria. This strong conclusion holds regardless of whether the differences in performance are statistically significant. We conclude that our theoretical findings are useful for model identification, and the tools we propose can be directly applied to investigate processes that exhibit structural changes.

Acknowledgement

The author would like to extend his gratitude to the editor of the journal and to the reviewers for their insightful comments and valuable suggestions.

References

Alemohammad, N., Rezakhah, S., and Alizadeh, S. H. (2020). Markov switching asymmetric GARCH model: stability and forecasting. *Statistical Papers*, **61**, 1309–1333.

Bauwens, L., Preminger, A. and Rombouts, J., V., K. (2010). Theory and inference for a Markov switching GARCH model. *Econometrics Journal*, **13**, 218–244.

Bibi, A. (2021). Asymptotic properties of QMLE for seasonal threshold GARCH model with periodic coefficients. *Statistical Methods and Applications*, **30**, 477–514.

Bibi, A., and Ghezal, A. (2015). On the Markov-switching bilinear processes: stationarity, higher-order moments and β-mixing. *Stochastics An International Journal of Probability and Stochastic Processes*, **87**, 919–945.

Bibi, A., and Hamdi, F. (2025). Covariance analysis and GMM estimation of Markov switching bilinear processes. *Statistics*, **59**, 354–387.

Boubacar M., Y., and Rabehasaina, L. (2020). Estimation of weak ARMA models with regime changes. *Statistical Inference for Stochastic Processes*, **23**, 1–52.

Brockwell, P. J. and Davis, R. A. (1987). *Time Series: Theory and Methods*. Springer-Verlag, New York.

Francq, C., and Zakoïan, J. M. (2001). Stationarity of multivariate Markov-switching ARMA models. *Journal of Econometrics*, **102**, 339–364.

Francq, C., and Zakoïan, J. M. (2005). L_2-structures of standard and switching-regime GARCH models. *Stochastic Processes and Their Applications*, **115**, 1557–1582.

Gabr, M. M. (1988). On the third-order moment structure and bispectral analysis of some bilinear time series. *Journal of Time Series Analysis*, **9**, 11–20.

Granger, C. W. J. and Anderson, A. (1978). *An Introduction to Bilinear Time Series Models.* Vandenhoeck and Ruprecht, Gottingen.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the Econometric Society*, **57**, 357–384.

Lee, O .(2005). Probabilistic properties of a nonlinear ARMA process with Markov switching. *Communications in Statistics: Theory and Methods*, **34**, 193–204.

Liu, J. (1992). On stationarity and asymptotic inference of bilinear time series models. *Statistica Sinica*, **2**, 479–494.

Marcucci, J. (2005). Forecasting stock market volatility with regime-switching GARCH models. *Studies in Nonlinear Dynamics and Econometrics*, **9**, 1–53.

Martin, C. M. (1999). A note on the third-order moments structure of a bilinear model. *Portugaliae Mathematica*, **16**, 115–125.

Stelzer, R. (2009). On Markov-switching ARMA processes-stationarity, existence of moments, and geometric ergodicity. *Econometric Theory*, **25**, 43–62.

Wee, D. C., Chen, F., and Dunsmuir, W. T. (2022). Likelihood inference for Markov switching GARCH (1, 1) models using sequential Monte Carlo. *Econometrics and Statistics*, **21**, 50–68.

Yang, M. (2000). Some properties of vector autoregressive processes with Markov-switching coefficients. *Econometric Theory*, **16**, 23–43.

Yao, J. F., and Attali, J. G. (2000). On stability of nonlinear AR processes with Markov switching. *Advances in Applied Probability*, **32**, 394–407.

Stochastic Models in Probability and Statistics
Vol. 2, No. 2, 2025, pp. 159–198. Research Paper

On the generalized last time buy problem: an application of pseudo-deterministic stopping times to end-of-life management

J. B. G. Frenk[†*], **Mustafa Hekimoğlu**[‡] **and Ayda Amniattalab**[§]

[†]*Industrial Engineering Department, Faculty of Engineering, Gebze Technical University, Kocaeli, Turkey*

[‡]*Department of Industrial Engineering, Kadir Has University, Istanbul, Turkey*

[§]*Department of Engineering and Natural Sciences, Sabancı University, Gebze, Turkey*

Email(s): frenk@gtu.edu.tr, mustafa.hekimoglu@khas.edu.tr, ayda@sabanciuniv.edu

Abstract. The End-of-life phase is the longest phase of the different life cycles of a product. This phase, characterized by a last time buy order, is known for increasing supply risk and decreasing demand rate. In this paper, starting initially with a repair replacement policy the manufacturer optimizes the order size of the last time buy and selects a time switching to another policy substituting failed components.The defective items stochastic arrival process is given by a non-homogenous Poisson process and the switch time is modeled as a stopping time of this demand process. Since the optimal switching time within the class of stopping times can be difficult to implement, we introduce in this paper the set of so-called pseudo-deterministic stopping times being the minimum of a deterministic stopping time and the time to depletion of the spare part inventory. We show that the optimal pseudo-deterministic stopping time satisfies some nice properties under realistic assumptions on the arrival rate function of the Poissonian defective items arrival process and the substitution cost function of the alternative policy. Using these properties an efficient algorithm is proposed to determine the optimal pseudo-deterministic stopping time. Although in general an optimal stopping time (optimal among the class of all stopping times) does not belong to the class of pseudo-deterministic stopping times, we numerically show and explain in the presence of high penalty costs that the objective value of the optimal pseudo-deterministic stopping time is close to the objective value of the optimal stopping time.

Keywords: End-of-life inventory problem; Martingales; Non-homogeneous Poisson process; Spare parts inventory management; Stopping time.

1 Introduction

Due to rapid technology developments the production time of newly introduced products in the market shortened considerable to sometimes only a few months. By these developments it became more difficult

*Corresponding author
Received: 22 February 2025 / Accepted: 30 October 2025
DOI: 10.22067/smps.2025.92293.1042

to deal with after sales services like for example warranty period requirements lasting for several years. At the same time, customers are also seeking support for their equipment even after the expiration date of their warranty, and they value manufacturers based on their aftersales service performance (Ahmad and Butt, 2012). Similarly, some customers are willing to pay extra for longer manufacturer care through extended service contracts (Padmanabhan and Rao, 1993). To respond to this demand and preserve their brand perception in the market, Original Equipment Manufacturers (OEMs) pay increasing attention to aftersale services of products which are at the end of their life cycle (Cohen et al., 2006).

Management of aftersales services for end-of-life products is problematic due to both demand and supply-side problems in spare parts supply chains. Demand-side problems in spare parts supply occur due to the uncertain spare parts demand over time as customers use only the product during a limited amount of time in the market and these products are subject to random failure times. At the supply side, OEMs need to deal with increasing risk of losing suppliers' support (Hekimoğlu et al., 2018; Li et al., 2016) for original spare parts mainly due to technological (Solomon et al., 2000) or economic reasons (Li et al., 2016).

OEMs utilize different strategies to deal with supply-side problems of original spare parts of a given product. The two most common strategies are recognized as last time buy and or development of a substitute component to guarantee the supply of spare parts (Shen and Willems, 2014). When a product at some OEM becomes end-of-life and is out of production, spare part manufacturers will eventually request the OEM to place a last order for original spare parts. For this order, which is referred to as the *Last Time Buy* order, OEMs need to consider the total amount of spare part demand until the planned date of End-of-Support (EoS), at which the OEM no longer provides repair services for that product.

The size of the Last Time Buy order is critical for OEMs as ordering too much original spare parts leads to obsolete spare parts, increasing salvage costs and economic losses. On the other hand, if the size of the Last Time Buy order is insufficient, this results in unsatisfied customer demand, possibly leading to the loss of goodwill and customers' brand loyalty (Padmanabhan and Rao, 1993). To safeguard against these problems an OEM can alternatively next to a repair-replacement policy also utilize a substitution-based policy. This means the company uses as a policy original spare parts in a repair-replacement policy if a defective product, which cannot be repaired, appears until a particular point in time before the End-of-Support date. After that time the OEM switches to an alternative substitution policy not depending on original spare parts. The combination of these two spare parts policies is recognized as *bridge buy* in the literature (Shen and Willems, 2014). We refer to this combined strategy as *Generalized Last Time Buy* since a switching time being equal to the End-of Support date leads to the classic Last Time Buy problem (Teunter and Fortuin, 1999). In the Generalized Last Time Buy problem, substitute products can be obtained by either taking over suppliers' production lines (Shen and Willems, 2014), using 3D printing (Westerweel et al., 2018) or finding an alternative supplier (Shi and Liu , 2020). Therefore, the OEM needs to make two critical decisions for the optimal solution of the Generalized Last Time Buy problem: At time 0, the OEM decides on the size of an Last Time Buy order and then chooses a switch-to-substitute time. The motivation of this study is to investigate how such a more general switching policy affects both the risk of obsolete spare parts and the supply-side risks of obtaining spare parts and how such a general switching policy increases the service level to customers. For a practical example in airline industry related to these risks the reader is referred to (Hekimoğlu et al. (2018)). Since the optimal switching policy is difficult to compute (Frenk et al. (2019b)) we are also motivated in this study in identifying a simple class of switching polices called the class of pseudo-deterministic policies which under certain conditions achieve expected objective costs close to the expected costs of the optimal switching policy among all stopping rules.

1.1 Problem setting

In this paper, we consider a problem setting where customers bring their defective products for repair to the OEM, which accepts them to its repair facility until the time of the End-of Support announcement.

It is assumed in all of the early and later literature (see for example (Behfard et al., 2015; Fortuin, 1980, 1981; Krikke and van Der Laan, 2011; Teunter and Fortuin, 1999)) that in a continuous time setting a non-homogeneous Poissonian defective items arrival process is a good representation of the demand for spare parts. If the arrival of a defective item is before the switch-to-substitute point, then the product's defective component is removed and sent to inspection for repair feasibility. If a repair is possible, the repaired component is installed back to the product and if the component cannot be repaired and there is available stock, the defective component is replaced with an original part to complete the repair service. If there is no original replacement in stock, the repair process uses a substitute part possible from an external source. This business process is depicted in Figure 1.

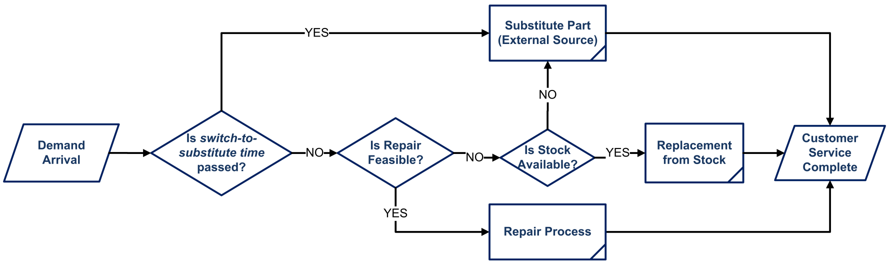

Figure 1: Repair process at the arrival of a defective product

The Generalized Last Time Buy strategy consists of the sequential utilization of a *repair-replacement* and a *repair-with-substitute* policy. This particular structure of the Generalized Last Time Buy strategy, which checks the switch-to-substitute point before checking the repairability of a part, allows OEMs to exploit decreasing substitution cost in time. We may justify this assumption since after the switching time to an alternative policy the repair center is closed and any defective product will be handled according to the new policy. The OEM now needs to make an optimal plan including the optimal size of a Last Time Buy order (at time 0) and the optimal switch-to-substitute time to keep their after sales services running. Note that our *repair-with-substitute* policy is dubbed *alternative* policy by Frenk et al. (2019b). Similar to Frenk et al. (2019a),Frenk et al. (2019b) we follow the same cost setup and our switch-to-substitute decision is formulated as an optimal stopping time problem. As such this paper can be seen as a review, extension and generalization of the results discussed in Frenk et al. (2019a) and Frenk et al. (2019b). To obtain the optimum solution of the Generalized Last Time Buy problem, we first derive the objective function for any switching time represented by an arbitrary stopping time following a simplified approach as done in Frenk et al. (2019a). Next, by substitution, we evaluate the objective function for the new defined subclass of switching times represented by the so-called *pseudo-deterministic stopping times*. These stopping times representing the switching time between different policies are defined as the minimum of a deterministic stopping time as discussed in detail in Frenk et al. (2019a) and the random time of inventory depletion of the Last Time Buy order received at time zero. According to the authors knowledge this is the first time the subclass of pseudo-deterministic switching times is discussed in the Generalized Last Time Buy problem literature and its mathematical properties are derived for both a piece-wise constant substitution cost function and a non-homogeneous Poisson defective items arrival process also having a piecewise constant arrival rate function. The practical and theoretical reason for studying this particular subclass of switching times between policies will be explained in the following paragraphs. First of all, we observe it is more difficult to compute the optimal pseudo-deterministic policy for a non-homogeneous Poisson defective items arrival process having an arbitrary bounded arrival rate function than for an arrival process having a piece-wise constant arrival rate function. In the later case the computation of this optimal pseudo-deterministic policy is very easy. Since any non-homogeneous Poisson process with a bounded arrival rate function can be approximated arbitrarily closely by a non-homogeneous Poisson process with a piece-wise constant arrival rate function it follows that the restriction to this class of arrival processes of defective items is not too restrictive. Our theoretical results and new algorithm for identifying the optimal pseudo-deterministic policy for this class of arrival processes having a piece-wise constant arrival rate function now complement and extend the results discussed in Frenk et al. (2019a), Frenk et al. (2019b) and Javadi (2018).

The motivation for studying this particular class of policies is that our numerical results suggest that the cost of using a switching time represented by the optimal *pseudo-deterministic stopping time* is almost the same as the cost of the optimal switching time represented by the optimal stopping time within the class of arbitrary stopping times. This indicates that the optimal *pseudo-deterministic stopping time* is a good approximation of the optimal stopping time within the class of arbitrary stopping times. The main reason for this empirical observation is the absence of penalty cost within the class of pseudo-deterministic policies. These penalty costs occur when both a defective non-repairable item arrives and at that time no inventory of original spare parts is available and we still apply the repair-replacement policy. If the penalty costs are high at the occurrence of these events any optimal policy tries to avoid these penalty costs by switching in time from the repair-replacement policy to an alternative policy and so such a policy acts like a pseudo-deterministic policy. At the same time, an optimal pseudo-deterministic stopping time is much easier to compute and due to its simplicity easier to implement in practice.

Our problem setting can be motivated by the end-of-life management of electronic products, whose manufacturers aim to provide post-warranty repair service to keep their customers loyal. Due to the rapid pace of technological change in the semi-conductor industry, component manufacturers frequently update their product assortment and shift their capacity to the components of newer models with higher profit potential. Component suppliers usually call for Last Time Buy order, forcing OEMs to decide on the order size of the last order and other supply alternatives.

This paper consists of 5 sections. In the remainder of this section, we briefly review the relevant literature and articulate the contribution of our study. Section 2 describes the mathematical model also discussed in Frenk et al. (2019a) and the different classes of policies and yields simplified proofs of already known results for this model. In Section 3, we derive under some reasonable assumptions properties of the class of pseudo-deterministic stopping times and at the same time give an easy procedure to compute the optimal policy within this class. These results seems to be new in the literature. Section 4 gives by means of easier proofs then the ones used in Frenk et al. (2019a) how to approximate the optimal policy within the class of all bounded stopping times for the generalized last buy decision problem or end of life decision problem. Finally Section 5 includes numerical results, while Section 6 concludes the paper.

1.2 Literature review

The relevant literature for our study mainly consists of the studies from end-of-life management of durable products.

The last stage of a capital product's life cycle starts with the end of the manufacturing of a product by OEMs. Some time later, component suppliers start announcing their end-of-support dates and ask OEMs to place their last orders. In the literature such an order is called a Last Time Buy order (Bradley and Guerrero, 2008, 2009; Fortuin, 1980, 1981; Frenk et al., 2019a; Teunter and Fortuin, 1999; Teunter and Haneveld, 1998). For determining the size of the Last Time Buy order, OEMs need to make an estimation on the total spare parts demand from their products in use. The first contributions in the literature on determining the size of a Last Time Buy order focused in a discrete time setting solely on using only a repair-replacement policy during the remaining economic life time of the product(Fortuin, 1980; Teunter and Fortuin, 1999). In a recent study, Hur et al. (2018) considers the size of a Last Time Buy order decision in a continuous time setting.

However, in many business cases, companies are interested in complementing decisions about the size of the Last Time Buy order with other policies to replace defective items such as removal of repairable parts from phased out products (Behfard et al., 2015; Frenk et al., 2019a; Van Kooten and Tan, 2009; Krikke and van Der Laan, 2011; Pourakbar et al., 2014), finding an alternative supplier of a substitute

(Frenk et al., 2019b; Pourakbar et al., 2012; Sahyouni et al., 2010) or redesigning the entire production line of components (Shen and Willems, 2014; Shi and Liu , 2020). In Krikke and van Der Laan (2011) a case study and a simple heuristic solution is provided for determining the size of a Last Time Buy order in combination with the availability of repairable spare parts that are removed from phased out products. They address the impact of phaseouts on spare parts demand and the availability of alternative supply sources. The same problem setting is utilized by Pourakbar et al. (2014) for the joint optimization of returned repairable parts from phaseouts and the Last Time Buy Order size. Shi (2019) extended this problem setting with actively collected end-of-life products. They explicitly model the dynamics of a installed base to jointly control the Last Time Buy order size, repairable parts and the size of the installed base. Design refresh options are another important tactic to complement Last Time Buy order decisions. Shi and Liu (2020) considers a problem setting where the Last Time Buy Order size, design refresh and end-of-support decision are jointly optimized in a discrete time setting with respect to the profit maximization criterion. In addition, Inderfurth and Kleber (2013), Inderfurth and Mukherjee (2008) and Bayındır et al. (2007) consider re-manufacturing options to satisfy incoming spare parts demand.

Another relevant research stream for our study is on spare parts management during the warranty period. (van der Heijden and Iskandar, 2013; Huang et al., 2007, 2008; Kim and Park, 2008; Sahyouni et al., 2010). van der Heijden and Iskandar (2013) consider joint optimization of the size of the Last Time Buy order and repair decisions for products with active warranty. They develop cost and service level formulations using a discrete time setting. Huang et al. (2007) consider a discrete-time problem setting where a manufacturer receives demand for new products and warranty claims of old ones. They show the optimality of a state-dependent base stock policy using Markov decision processes. This problem setting is extended with age-dependent warranty claims in the installed base by Huang et al. (2008). Kim and Park (2008) employ continuous time optimal control theory to optimize pricing and production of durable products by considering their spare parts costs during warranty periods. They explicitly model the positive impact of a warranty period on the products' demand. Sahyouni et al. (2010) study the joint optimization of repair and Last Time Buy order quantity for a deterministic repair demand from products under warranty in a continuous time setting. In their model, which is more suited to products having a short life cycle, they focus on the joint optimization of the Last Time Buy order size and end-of-repair point after which failed products are replaced by substitute products.

1.3 Our contribution to the literature

Our paper contributes to the existing literature of the Generalized Last Time Buy decision problem by analyzing in detail a subclass of switching times from a repair-replacement policy to an alternative policy in the optimal stopping formulation of the Generalized Last Time Buy decision problem as discussed in Frenk et al. (2019a). It is shown that this subclass of switching times satisfies desirable properties under very general assumptions on the non-homogeneous Poisson arrival process of defective items and the substitution cost function of an alternative policy. The optimal policy within this class of these so-called pseudo-deterministic policies can be easily computed and can be used as a heuristic solution replacing the optimal, more complicated, optimal policy within the class of all arbitrary bounded stopping times. It is verified and also explained in the computational section that under certain conditions this can be done without a significant loss in the objective value. At the same time, the optimal policy within the class of pseudo-deterministic policies has a clear interpretation contrary to the optimal policy within the class of all bounded stopping times.

The closest studies to our paper are Frenk et al. (2019a); Javadi (2018); Pourakbar et al. (2012); Shi and Liu (2020). In Frenk et al. (2019a), which extends Pourakbar et al. (2012), the optimal solution of this problem is studied and solved by a dynamic programming algorithm. In Shi and Liu (2020) a similar problem is addressed in a discrete time setting. Under deterministic demand assumptions the same problem is solved in Sahyouni et al. (2010) using mathematical programming techniques. To the best of our knowledge, this paper is the first one addressing the use of pseudo-deterministic policies in the joint optimization of the size of the Last Time Buy order and switching time to an alternative policy in a continuous time setting for a non-homogeneous Poisson defective items arrival process.

2 On the generalized last time buy problem

In this section we introduce in subsection 2.1 the random arrival process of defective items. In subsection 2.2 we define the costs parameters of the generalized last time buy problem, while in section 2.3 we introduce the objective function for this problem. In subsection 2.4 we formulate the associated minimization problem and derive in subsection 2.5 a lowerbound on the optimal objective value. This is needed to derive an upperbound on the optimal Last Time Buy order size as expressed in Lemma 3 for nonnegative scrapping values and in Lemma 5 for arbitrary scrapping values. Finally in subsection 2.6 we derive some qualitative results for the new class of pseudo-deterministic stopping rules. As such these six subsections reviews, simplifies and extend and complements the model and results in Frenk et al. (2019a) and Frenk et al. (2019b).

2.1 The arrival process of the defective items process

To formulate our end-of-life inventory or generalized last time buy problem, we assume that defective products arrive according to a non-homogeneous Poisson point process for a repair or replacement. To introduce this arrival process let $(\Omega,\mathscr{H},\mathbb{P})$ be a probability space hosting the point process $(T_i,R_i)_{i\in\mathbb{N}}$. The random variable T_i, $i\in\mathbb{N}$ denotes the arrival time of the ith customer having a defective product and requesting repair. The counting process of defective products $N\equiv\{N(t):t\geq 0\}$ is defined by

$$N(t):=\textstyle\sum_{i=1}^{\infty}1_{\{T_i\leq t\}},t\geq 0, \tag{1}$$

and assumed to be a non-homogeneous Poisson process with a bounded Borel arrival rate function $\lambda(.)$. The random variables R_i, $i\in\mathbb{N}$, on the other hand, are independent and identically distributed Bernoulli random variables indicating the condition of the defective items. They are defined as

$$R_i=\begin{cases}1 & \text{if the product can be repaired}\\ 0 & \text{otherwise}\end{cases}$$

with probability $0\leq q=\mathbb{P}(R_i=1)\leq 1$. The thinned arrival processes $N_0:=\{N_0(t):t\geq 0\}$ and $N_1=\{N_1(t):t\geq 0\}$ given by

$$N_0(t):=\textstyle\sum_{i=1}^{\infty}(1-R_i)1_{\{T_i\leq t\}},N_1(t):=\sum_{i=1}^{\infty}R_i1_{\{T_i\leq t\}}, \tag{2}$$

count the number of non-repairable and repairable products arriving over time. It is well known that the arrival processes N_0 and N_1 are independent non-homogeneous Poisson processes with arrival rate functions $\lambda_0(.)=(1-q)\lambda(.)$, $\lambda_1(.)=q\lambda(.)$ respectively (cf. Çınlar (2011)) having mean arrival functions $\Lambda_0(.)=(1-q)\Lambda(.)$ and $\Lambda_1(.)=q\Lambda(.)$ with $\Lambda(t)=\int_0^t\lambda(u)du$ for $0\leq t\leq T$. In the sequel, we let $\mathbb{F}\equiv(\mathscr{F}_t)_{t\geq 0}\subseteq\mathscr{H}$ denote the filtration of the point process $(T_i,R_i)_{i\in\mathbb{N}}$; that is, the flow of information associated with both the arrival times of the products and their condition.

Let T denote the end of service time at which the OEMs' service obligations with respect to the product expires. For the optimal application of the Generalised Last Time Buy policy, the manufacturer makes two decisions: the size of the Last Time Buy order at time 0, $x\in\mathbb{Z}_+=\{0,1,2,\ldots\}$, and the optimal stopping time $\tau\in[0,T]$ of switching from a classical repair-replacement policy to an alternative policy. This is a (possibly random) stopping time belonging to the set of all bounded stopping times with respect to the filtration $\mathbb{F}$. The subset of deterministic stopping times is denoted by $\mathbb{F}_0$. Recall that for the deterministic stopping time $\tau\overset{a.s}{=}T$ the Generalised Last Time Buy optimization problem reduces to the Last Time Buy optimization problem in which we only need to determine the size of the Last Time Buy order.

For the decision making problem, we consider the finite horizon, continuous compounding discounted cost criteria with discount rate $\delta \geq 0$. Figure 1 depicts the control process under the Generalized Last Time Buy policy. The total cost of these two decisions for a given (x,τ)-policy consists of cost components related with acquisition, inventory holding, repair and replacement. These cost parameters are explained in the next subsection and are the same as discussed in Frenk et al. (2019a).

2.2 The cost parameters of the generalized last time buy problem

Acquisition of spare parts take place at time 0 and parts are delivered immediately. The acquisition cost function is denoted by $c: \mathbb{Z}_+ \mapsto \mathbb{R}_+$, where $c(0) = 0$. Also the holding cost rate for the delivered parts is $h > 0$ per spare part per unit of time.

The used policy up to the end of service time T is now summarized by the following procedure. Starting with x units representing the size of the Last Time Buy order, the OEM utilizes the *repair-replacement* policy until some random stopping time $\tau \leq T$. Under this repair-replacement policy, if a repair is feasible for an arriving defective component, it is always repaired at some repair cost c_{re} plus some service cost c_{se}. If the component is beyond repair and the inventory level of spare parts is non-zero, the defective component in the item is replaced with a spare part from the inventory at service cost c_{se}. If no spare part is available in inventory, the beyond repairable component is replaced with a substitution spare part supplied from an external source like a gray market or a 3D printer (Figure 1). The cost of this substitution is given by the right continuous function $c_a : [0,T] \to \mathbb{R}^+$. Utilizing the substitution policy during the time one should use the repair-replacement policy also leads to an additional penalty cost function $p : [0,\tau] \to \mathbb{R}^+$. This penalty cost function can be motivated by customers' loss-of-goodwill or transportation of the substitute part from another location. Therefore, the total penalty cost of being forced to use the alternative substitution policy at time $0 \leq t \leq \tau$ during the time one should use the repair-replacement policy is given by

$$c_{ap}(t) := c_a(t) + p(t). \tag{3}$$

We assume that $c_{ap}(t) \geq c_{se}$ for $0 \leq t \leq \tau$ to avoid trivial domination of substitution parts over spare parts available in inventory. It is assumed that the functions c_a and p are both non-increasing and right continuous. Namely, the substitute component gets cheaper over time thanks to technological progress in production processes and the penalty of using a substitute component under the repair-replacement policy is a decreasing function of time.

After time $\tau \leq t \leq T$ until the end of service time T, the OEM first discards at time τ the existing inventory at a scrapping cost of c_{scr} per item and abandons after this time for any defective item arriving at time t the *repair-replacement* policy and replaces it by the substitution policy at the substitution cost $c_a(t)$. All costs of the repair services are indicated next to the associated processes in Figure 1. In our formulation, all cost terms are positive except the scrapping value c_{scr}, which is allowed to be negative. This means there can be a net revenue associated with these scrapped parts. To avoid pathological cases where ordering is profitable because of scrapping, we assume in this study that the function $x \mapsto c(x) - c_{scr}^- x$ is increasing with $c_{scr}^- := -\min\{c_{scr}, 0\}$ and

$$\lim_{x\uparrow\infty} c(x) - c_{scr}^- x = \infty. \tag{4}$$

This implies for the special case that the scrapping costs are positive, then the acquisition cost function c is increasing and satisfies $\lim_{\uparrow\infty} c(x) = \infty$. Also for scrapping costs negative this implies that the function $x \to c(x) + c_{scr} x$ is increasing. For the scrapping action to be economically justifiable, we must impose the condition that $h - \delta c_{scr} \geq 0$. If this condition fails to hold, instead of scrapping an item at some time τ, we can keep it indefinitely in inventory at a total cost of $h \int_\tau^\infty e^{-\delta u} du = (h/\delta) e^{-\delta\tau}$ which would be less than $c_{scr} e^{-\delta\tau}$. These three conditions on the cost functions and the parameters listed in the previous relations always hold in this study unless stated otherwise.

2.3 The objective function of the generalized last time buy problem

In the decision making problem under consideration, a Last Time Buy order of size x and switching time $\tau \leq T$ are determined, which we refer to as a (x,τ)-policy. The first variable x is static, and its value is determined at time zero. The switching decision, on the other hand, can be dynamic, and in the general formulation of the problem τ is a stopping time of the filtration $\mathbb{F}$.

In this section, we derive the expected discounted cost $C(x,\tau)$ of any (x,τ)-policy, $x \in \mathbb{Z}_+$, $\tau \in \mathbb{F}$ and introduce the optimization problem to be solved. To this end, we introduce the random variable $\sigma_x := \inf\{t > 0 : N_0(t) \geq x\}$, denoting the (random) time of inventory depletion for an order size of x. Observe the total expected discounted cost is the sum of the procurement and expected discounted operation costs. The procurement cost is given by $c(x)$. To derive the expected discounted operation costs we utilize the following results about martingales. For any locally bounded Borel measurable function f and any non-homogeneous Poisson process N with locally bounded Borel arrival rate function μ defined on the probability space $(\Omega, \mathscr{H}, \mathbb{P})$ it follows that the stochastic process $X = \{X(t) : t \geq 0\}$, given by $X(t) := \int_0^t f(s)dN(s) - \int_0^t f(s)\mu(s)ds$ for $t \geq 0$, is a right-continuous martingale (cf. Çınlar (2011)). For these so-called Poissonian martingales and any bounded stopping time τ it holds that $\mathbb{E}(X(\tau)) = 0$. This means

$$\mathbb{E}\left(\int_0^\tau f(t)dN(t)\right) = \mathbb{E}\left(\int_0^\tau f(t)\mu(t)dt\right). \tag{5}$$

This is known as Doob's stopping theorem (cf.Çınlar (2011)) and we will make frequently use of this result in the computation of the expected discounted operation costs. The expected discounted operation costs consist of the following components for any (x,τ)-policy with $x \in \mathbb{Z}_+$ and $\tau \in \mathbb{F}$:

- **Inventory holding costs:** We switch to the repair-with-substitute policy at time $\tau \leq T$ and scrap at that time (possibly) leftover inventory of spare parts. Hence, the random discounted inventory holding costs are given by $h\int_0^\tau e^{-\delta u}(x - N_0(u))^+ du$ with $(z)^+ := \max(z,0)$. This shows that the expected discounted holding costs $C_{inv}(x,\tau)$ of any (x,τ)-policy equal

$$C_{inv}(x,\tau) = h\mathbb{E}\left(\int_0^\tau e^{-\delta u}(x - N_0(u))^+ du\right). \tag{6}$$

- **Service costs:** Service costs arise during the cost of running repair operations for both repairable and non-repairable products. For a repairable product service costs occur during the repair-replacement policy from time 0 to time τ in any (x,τ)-policy. For a non-repairable product, service costs occur in case of positive spare parts inventory at the arrival time of a defective product. Hence, for non-repairable products, service costs only need to be paid from time 0 up to time $\tau \wedge \sigma_x := \min\{\tau, \sigma_x\}$. This shows that the random discounted service costs are given by

$$c_{se}\int_0^\tau e^{-\delta u} dN_1(u) + c_{se}\int_0^{\tau\wedge\sigma_x} e^{-\delta u} dN_0(u).$$

  Applying now relation (5) for properly chosen functions f and using both point processes N_0 and N_1 it follows for the bounded stopping time $\tau \leq T$ that the expected discounted service costs $C_{se}(x,\tau)$ of any (x,τ)-policy equal

$$\begin{aligned} C_{se}(x,\tau) &= c_{se}\mathbb{E}\left(\int_0^\tau e^{-\delta u}\lambda_1(u)du\right) + c_{se}\mathbb{E}\left(\int_0^{\tau\wedge\sigma_x} e^{-\delta u}\lambda_0(u)du\right) \\ &= qc_{se}\mathbb{E}\left(\int_0^\tau e^{-\delta u}\lambda(u)du\right) + (1-q)c_{se}\mathbb{E}\left(\int_0^{\tau\wedge\sigma_x} e^{-\delta u}\lambda(u)du\right). \end{aligned} \tag{7}$$

- **Repair costs:** We incur repair costs during the repair-replacement phase from time 0 to τ. Hence the random discounted repair costs are given by $c_{re}\int_0^\tau e^{-\delta u}dN_1(u)$. Applying the same arguments as for service costs the expected discounted repair costs $C_{re}(x,\tau)$ of any (x,τ)-policy equal to

$$C_{re}(x,\tau) = c_{re}\mathbb{E}\left(\int_0^\tau e^{-\delta u}\lambda_1(u)du\right) = qc_{re}\mathbb{E}\left(\int_0^\tau e^{-\delta u}\lambda(u)du\right). \tag{8}$$

- **Substitution costs:** The random discounted costs of applying the repair-with-substitute policy consist of the substitution cost before and after time τ. The random discounted cost of using the alternative policy is given by

$$\int_\tau^T e^{-\delta u}c_a(u)dN(u) + \int_{\tau\wedge\sigma_x}^\tau e^{-\delta u}c_{ap}(u)dN_0(u).$$

Applying the same arguments as before and using relation (3) the expected discounted alternative policy costs $C_a(x,\tau)$ of any (x,τ)-policy are as follows:

$$\begin{aligned}
C_a(x,\tau) &= \mathbb{E}\left(\int_\tau^T e^{-\delta u}c_a(u)\lambda(u)du\right) + (1-q)\mathbb{E}\left(\int_{\tau\wedge\sigma_x}^\tau e^{-\delta u}c_{ap}(u)\lambda(u)du\right) \\
&= \begin{cases} \int_0^T e^{-\delta u}c_a(u)\lambda(u)du + \mathbb{E}(\int_0^\tau e^{-\delta u}[(1-q)c_{ap}(u)-c_a(u)]\lambda(u)du \\ -(1-q)\mathbb{E}\left(\int_0^{\tau\wedge\sigma_x} e^{-\delta u}c_{ap}(u)\lambda(u)du\right) \end{cases} \\
&= \begin{cases} \int_0^T e^{-\delta u}c_a(u)\lambda(u)du + \mathbb{E}(\int_0^\tau e^{-\delta u}[(1-q)p(u)-qc_a(u)]\lambda(u)du \\ -(1-q)\mathbb{E}\left(\int_0^{\tau\wedge\sigma_x} e^{-\delta u}c_{ap}(u)\lambda(u)du\right). \end{cases}
\end{aligned} \tag{9}$$

- **Scrapping costs:** The random discounted scrapping costs at time τ are given by $c_{scr}e^{-\delta\tau}(x-N_0(\tau))^+$. This shows that the expected discounted scrapping costs $C_{scr}(x,\tau)$ of any (x,τ) policy equal

$$C_{scr}(x,\tau) = c_{scr}\mathbb{E}(e^{-\delta\tau}(x-N_0(\tau))^+). \tag{10}$$

Adding up the separate operational costs derived in relations (6)-(10), the expected discounted operation cost $C(x,\tau)$ of any (x,τ)-policy is given by

$$C(x,\tau) = \begin{cases} h\mathbb{E}(\int_0^\tau e^{-\delta u}(x-N_0(u))^+du) + c_{scr}\mathbb{E}(e^{-\delta\tau}(x-N_0(\tau))^+) \\ +\mathbb{E}\left(\int_0^\tau e^{-\delta u}[q(c_{re}+c_{se}-c_a(u))+(1-q)p(u)]\lambda(u)du\right) \\ +(1-q)\mathbb{E}\left(\int_0^{\tau\wedge\sigma_x} e^{-\delta u}[c_{se}-c_{ap}(u)]\lambda(u)du\right) + \int_0^T e^{-\delta u}c_a(u)\lambda(u)du. \end{cases} \tag{11}$$

To rewrite the expression in (11) in a more suitable form, we remind that for any first order stochastic processes $Y = \{Y(t), t\geq 0\}$ and $Z = \{Z(t) : t \geq 0\}$ it holds that

$$Y(t)Z(t) = Y(0)Z(0) + \int_0^t Y(u)dZ(u) + \int_0^t Z(u)dY(u), t > 0$$

This shows for any $0 \leq q \leq 1$ using $N_0(0) = 0$

$$\begin{aligned}
e^{-\delta\tau}(x-N_0(\tau))^+ &= e^{-\delta(\tau\wedge\sigma_x)}(x-N_0(\tau\wedge\sigma_x)) \\
&= x - \delta\int_0^{\tau\wedge\sigma_x} e^{-\delta u}(x-N_0(u))du - \int_0^{\tau\wedge\sigma_x} e^{-\delta u}dN_0(u)
\end{aligned}$$

and so by Doob's stopping theorem

$$\mathbb{E}(e^{-\delta\tau}(x-N_0(\tau))^+) = x - \delta\mathbb{E}\left(\int_0^\tau e^{-\delta u}(x-N_0(u))^+ du\right) - (1-q)\mathbb{E}\left(\int_0^{\tau\wedge\sigma_x} e^{-\delta u}\lambda(u)du\right). \quad (12)$$

Replacing the expectation for the scrapping value in (11) with the expression in (12) and rearranging some terms, we obtain the following more suitable representation of the expected discounted operation costs.

$$C(x,\tau) = \begin{cases} c_{scr}x + (1-q)\mathbb{E}\left(\int_0^{\tau\wedge\sigma_x} e^{-\delta u}\lambda(u)[c_{se}-c_{scr}-c_{ap}(u)]du\right) \\ +\mathbb{E}\left(\int_0^\tau e^{-\delta u}\lambda(u)[q(c_{se}+c_{re}-c_a(u))+(1-q)p(u)]du\right) \\ +(h-\delta c_{scr})\mathbb{E}\left(\int_0^\tau e^{-\delta u}(x-N_0(u))^+)du\right) + \int_0^T e^{-\delta u}\lambda(u)c_a(u)du. \end{cases} \quad (13)$$

In case we consider the subclass of policies $(x,\tau\wedge\sigma_x), \tau\in\mathbb{F}$ this yields by relation (13)

$$C(x,\tau\wedge\sigma_x) = \begin{cases} c_{scr}x + \mathbb{E}\left(\int_0^{\tau\wedge\sigma_x} e^{-\delta u}\lambda(u)[c_{se}+qc_{re}-(1-q)c_{scr}-c_a(u)]du\right) \\ +(h-\delta c_{scr})\mathbb{E}\left(\int_0^\tau e^{-\delta u}(x-N_0(u))^+ du\right) + \int_0^T e^{-\delta u}\lambda(u)c_a(u)du \end{cases} \quad (14)$$

and so

$$C(x,\tau) - C(x,\tau\wedge\sigma_x) = \mathbb{E}\left(\int_{\tau\wedge\sigma_x}^\tau e^{-\delta u}[q(c_{se}+c_{re}-c_a(u))+(1-q)p(u)]du\right). \quad (15)$$

In the following lemma, we use relation (15) to show that under certain conditions on the penalty and the substitution cost functions it is always optimal within the class of all stopping times to switch to the substitution policy before or at the moment the inventory level of spare parts is depleted. In Section 2.6 we will analyze this special class of policies with τ a constant.

Lemma 1. *If*

$$\inf_{0\leq t\leq T}\{q(c_{se}+c_{re}-c_a(t))+(1-q)p(t)\} \geq 0, \quad (16)$$

then the class of $(x,\tau\wedge\sigma_x)$-policies with $\tau\in\mathbb{F}$ contains an optimal policy.

Proof. Apply relation (15). □

If no item is repairable ($q=0$!), the (sufficient) condition of Lemma 1 is satisfied. In this case one can find an optimal policy belonging to this special class of policies. In many service systems penalty costs are very high. Hence the condition in Lemma 1 is mostly satisfied in practice especially when q is close to 0. Since the subclass of $(x,\tau\wedge\sigma_x)$-policies avoid the (possibly) high penalty costs occurring in practice, it is therefore worthwhile to study them in detail. We will focus on this set of policies with $\tau\in\mathbb{F}_0$ in subsection 2.6 and section 3. However, it might not always happen that a $(x,\tau\wedge\sigma_x)$-policy is optimal within the class of all bounded stopping time policies. In case of high substitution cost $c_a(.)$ being larger then $c_{se}+c_{re}$ and low penalty costs $p(.)$, and q close to one it might be cheaper (see proof Lemma 1!) to continue with the repair-replacement policy after the inventory level hits zero.

2.4 On the formulation of the generalized last time buy problem

Applying relation (13) the objective function of using a given (x,τ) policy consists of the summation of the procurement cost $c(x)$ and the operating cost $C(x,\tau)$. The corresponding last time buy optimization problem is then given by

$$\upsilon(P) = \inf_{x\in\mathbb{Z}_+,\tau\in\mathbb{F},0\leq\tau\leq T}\{c(x)+C(x,\tau)\}. \quad (17)$$

Hence we need to determine the parameters of a (x,τ)-policy, if it exists, attaining the infimum in the above optimization problem.

There is one instance of optimization problem (17) which can be solved easily. In case all the defective items are repairable, i.e. $q=1$, the stochastic counting process N_0 of non-repairable defective components listed in relation (2) becomes the zero process. Since for $q=1$ all items can be repaired without using spare parts, it is easy to check applying relation (13) it is optimal not to order any spare parts. Also, since the function c_a is decreasing and every time we use for a (repairable) defective item the repair decision against cost $c_{se}+c_{re}$ it is optimal to switch to the substitution policy at the earliest time $t \leq T$ satisfying $c_a(t) \leq c_{se}+c_{re}$. This result is formally stated in the next lemma.

Lemma 2. *If all defective items can be repaired and the three conditions on the parameters introduced in subsection 2.2 hold, then the optimal policy is not to order any spare parts at time* 0 *and the optimal switching time to the repair-with-substitute policy is given by* $\tau = \zeta_1$ *with*

$$\zeta_q = \inf\{0 \leq u \leq T : c_{se} + qc_{re} - c_a(u) \geq 0\}, 0 \leq q \leq 1, \tag{18}$$

with the convention $\inf\{\varnothing\} = T$ *and* $\varnothing$ *denoting the empty set. Also the optimal objective value* $\upsilon(P)$ *equals*

$$\upsilon(P) = \int_0^T e^{-\delta u}\lambda(u)\min\{c_{se}+c_{re}, c_a(u)\}du.$$

Proof. See Appendix. □

Since in Lemma 2 we know for the optimal policy that the time to switch to the alternative policy is already known at time zero this optimal policy is a so-called *static* policy. Hence by Lemma 2 we only need to consider in the remainder of this paper the optimization problem (17) satisfying $0 \leq q < 1$. To solve optimization problem (17) we first compactify the decision space by deriving an upper bound on the optimal order quantity. An easy upper bound valid for $c_{scr} \geq 0$ is given by the following result. Since the function $x \to \inf_{\tau\in\mathbb{F},0\leq\tau\leq T}\{c(x)+C(x,\tau)\}$ does not satisfy in general discrete convexity type properties this restriction to a finite number of possible order sizes is very useful in the computational section. Observe, for the general case of both either positive or negative values of c_{scr} one can show under an additional condition on the function c_{ap} an improved upper bound and this is shown in Lemma 6.

Lemma 3. *If* $c_{scr} \geq 0$ *and the three conditions on the parameters introduced in subsection 2.2 hold and* $x_U := \min\left\{x \in \mathbb{Z}_+ : c(x) > \int_0^T e^{-\delta u}\lambda(u)c_a(u)du\right\}$ *then an optimal order quantity* x_* *of the optimization problem (17) exists and it satisfies* $x_* \leq x_U$.

Proof. See Appendix. □

Lemma 3 implies that solving optimization problem (17) is equivalent to solving

$$\upsilon(P) = \min_{x\leq x_U, x\in\mathbb{Z}_+}\inf_{\tau\in\mathbb{F},0\leq\tau\leq T}\{c(x)+C(x,\tau)\}, \tag{19}$$

and this problem can be approximated by a *computable optimal stopping problem*. In Frenk et al. (2019a) an approximate solution to this problem is provided by replacing the set of all stopping times with the set of stopping times attaining only values from a finite subset of [0,T]. To bound the approximation error replacing the set of stopping times by the smaller set of stopping times only attaining values at a finite subset of $[0,T]$ we need a lower bound on $\upsilon(P)$. In the following subsection we provide under some additional assumption on the function $c_{ap}(.)$ an improved upperbound on the optimal order quantity x valid for any scrapping value c_{scr}. This improves the representation in (19).

2.5 An upper bound on the optimal order quantity for arbitrary scrapping values

In this subsection we derive in Lemma 6 an upper bound on the optimal order quantity of the generalized last buy decision minimization problem for arbitrary scrapping values c_{scr} under general conditions on the function $c_{ap}(.)$. To do so we first derive a lower bound on $\upsilon(P)$ in Lemma 5. To start our analysis introduce the function $L: \mathbb{Z}_+ \times \mathbb{F} \to \mathbb{R}$ given by

$$L(x,\tau) := \begin{cases} c(x) + h\mathbb{E}\left(\int_0^\tau e^{-\delta u}(x - N_0(u))^+ du\right) + c_{scr}\mathbb{E}(e^{-\delta\tau}(x - N_0(\tau))^+) \\ +\mathbb{E}\left(\int_0^\tau e^{-\delta u}[c_{se} + qc_{re} - c_a(u)]\lambda(u)du\right) + \int_0^T e^{-\delta u} c_a(u)\lambda(u)du. \end{cases} \tag{20}$$

If $c_{ap}(u) = c_{se}$ for every $0 \leq u \leq T$ it follows that the costs of the different events given by either an arrival moment of a non-repairable defective item at which one still uses the repair-replacement policy but the inventory level of spare parts is zero or the cost of replacement at a positive spare parts inventory level are the same. If this holds we obtain using $p(u) = c_{ap}(u) - c_a(u) = c_{se} - c_a(u)$ that for every $0 \leq u \leq T$

$$q(c_{re} + c_{se} - c_a(u)) + (1-q)p(u) = c_{se} + qc_{re} - c_a(u)$$

This implies by relation (11) that for this selection of the cost parameters the value $L(x,\tau)$ is the sum of the procurement and the operational costs for every $0 \leq t \leq T$ and so the result in Lemma 4 should not come as a surprise. Also by relation (11) it is easy to check that

$$c(x) + C(x,\tau) - L(x,\tau) = (1-q)\mathbb{E}\left(\int_{\tau\wedge\sigma_x}^{\tau} e^{-\delta u}\lambda(u)[c_{ap}(u) - c_{se}]du\right), \tag{21}$$

and by relation (12)

$$L(x,\tau) = \begin{cases} c(x) + c_{scr}x + (h - \delta c_{scr})\mathbb{E}\left(\int_0^\tau e^{-\delta u}(x - N_0(u))^+ du\right) + \int_0^T e^{-\delta u}\lambda(u)c_a(u)du \\ -c_{scr}(1-q)\mathbb{E}\left(\int_0^{\tau\wedge\sigma_x} e^{-\delta u}\lambda(u)du\right) + \mathbb{E}\left(\int_0^\tau e^{-\delta u}[c_{se} + qc_{re} - c_a(u)]\lambda(u)du\right). \end{cases} \tag{22}$$

The following result, which immediately follows from relations (21) and (22), is also shown in Frenk et al. (2019a) using a much more complicated proof.

Lemma 4. *If the three conditions on the parameters introduced in subsection 2.2 hold, then for every $\tau \in \mathbb{F}$ the function $x \to L(x,\tau)$ is increasing and $\lim_{x\uparrow\infty} L(x,\tau) = \infty$. Also, if additionally $c_{ap}(u) \geq c_{se}$ for every $0 \leq u \leq T$, then $c(x) + C(x,\tau) \geq L(x,\tau)$ for every (x,τ) policy.*

Proof. By relation (22) the first part follows, while the second part is an immediate consequence of relation (21). □

Clearly by the interpretation of c_{ap} being the cost of being forced to apply the substitution policy to a non-repairale defective item at a time one still applies the repair-replacement policy it is natural to assume that this cost is higher then the costs c_{se} representing the cost of applying the repair-replacement policy to that same item under expected conditions. Using Lemma 4 one can derive for every positive or negative scrapping value c_{scr} per item a positive lower bound on $\upsilon(P)$.

Lemma 5. *If the three conditions on the parameters introduced in subsection 2.2 hold and $c_{ap}(u) \geq c_{se}$ for every $0 \leq u \leq T$, then the optimal objective value $\upsilon(P)$ of the optimization problem (17) satisfies*

$$\upsilon(P) \geq \int_0^T e^{-\delta u}\min\{c_{se} + qc_{re}, c_a(u)\}\lambda(u)du > 0. \tag{23}$$

Proof. see Appendix. □

Using Lemma 4 and the natural condition $c_{ap}(u) \geq c_{se}$ for every $0 \leq u \leq T$ it is possible to derive a tighter upper bound $\bar{x}_U \leq x_U$ on the optimal order quantity than the one presented in Lemma 3. Contrary to the (weaker) one in Lemma 3 this upper bound holds for both positive and negative values of the scrapping value c_{scr}.

Lemma 6. *If the three conditions on the parameters introduced in subsection 2.2 hold and $c_{ap}(u) \geq c_{se}$ for every $0 \leq u \leq T$ and*

$$\bar{x}_U = \min\{x \in \mathbb{N} : c(x) + \min\{c_{scr}, 0\}x > \int_0^T e^{-\delta u}\lambda(u)\max\{0, c_a(u) - c_{se} - qc_{re}\}du\}, \tag{24}$$

then an optimal order quantity x_* of the optimization problem (17) exists and it satisfies $x_* \leq \bar{x}_U$.

Proof. See Appendix. □

Observe the upper bound $\bar{x}_U$ is easy to compute by bisection since by our assumptions the function $x \to c(x) + \min\{c_{scr}, 0\}x$ is increasing having limit infinity at infinity. By using the proof of Lemma 6, one can construct an upper bound knowing the value $c(\bar{x}) + C(\bar{x}, \bar{\tau})$ of the objective function for a given $(\bar{x}, \bar{\tau})$-policy. In this case the upper bound $\bar{x}_U(\bar{x}, \bar{\tau})$ is given by

$$\bar{x}_U(\bar{x}, \bar{\tau}) = \min\left\{x \in \mathbb{N} : c(x) + \min\{c_{scr}, 0\}x > c(\bar{x}) + C(\bar{x}, \bar{\tau}) - \int_0^T e^{-\delta u}\lambda(u)\min\{c_{se} + qc_{re}, c_a(u)\}du\right\}. \tag{25}$$

Clearly by Lemma 5 the value

$$c(\bar{x}) + C(\bar{x}, \bar{\tau}) - \int_0^T e^{-\delta u}\lambda(u)\min\{c_{se} + qc_{re}, c_a(u)\}du,$$

is non-negative and the lower the values of this difference leads to tighter bounds.

2.6 On the subclass of deterministic and pseudo-deterministic policies

Another important class of policies is the class of (x, τ) policies with $\tau \in \mathbb{F}_0$. These so-called *deterministic* policies are studied in detail in Frenk et al. (2019b). Observe these polices are sometimes also called *static* due to the following advantage. The decision maker knows applying those policies already at time 0 the switching time of the repair-replacement policy to the alternative policy and so this policy is easy-to-implement in practice. Not knowing the switching time at time zero, but avoiding possibly high penalty costs instead, leads us to study the class of $(x, \tau \wedge \sigma_x)$ policies with $\tau \in \mathbb{F}$. For arbitrary bounded stopping times $\tau \in \mathbb{F}$ it was shown in Lemma 1 that under some reasonable conditions the class of $(x, \tau \wedge \sigma_x)$ polices indeed contains an optimal one among all (x, τ) policies, $x \in \mathbb{Z}_+$, $\tau \in \mathbb{F}$. We will now study this class of policies in more detail for any $\tau \in \mathbb{F}_0$ and call this class the class of *pseudo-deterministic* policies. These policies are not discussed previously in the literature and are still relatively easy to implement in practice as the decision maker switches from the repair-replacement policy to the alternative policy either at the deterministic time $\tau \in \mathbb{F}_0$ or at the random time σ_x of inventory depletion, whichever occurs first. Since in general penalty cost can be very high, it seems natural to assume that any optimal policy tries to avoid penalty cost or at least tries to minimize the probability of penalty occurrences. Hence it is useful to study this new class of policies. We will compute for different scenarios in Section 5 the probability of occurrence of a penalty and the gap between the cost of the optimal pseudo-deterministic policy and $\upsilon(P)$. In the next section, we will also show that it is extremely simple to compute an optimal pseudo-deterministic policy for the important class of piece-wise constant arrival rate and substitution policy costs functions. For these cost and arrival settings one can compute exactly the optimal policy within the class of pseudco deterministic policies instead of approximating the optimal policy. At the same time these cost and arrival settings approximate by any accuracy the general arrival and cost settings.

For the subset of deterministic and pseudo-deterministic policies, one can simplify considerably the expression for $C(x,\tau)$ in relation (13). By Fubini's theorem it follows for any Borel measurable function $f, \tau \in \mathbb{F}_0$ and $x \in \mathbb{Z}_+$ that

$$\mathbb{E}\left(\int_0^{\tau\wedge\sigma_x} f(u)du\right) = \mathbb{E}\left(\int_0^{\tau} f(u)1_{\{\sigma_x > u\}}du\right) = \int_0^{\tau} f(u)\mathbb{P}(N_0(u) < x)du.$$

This relation and relation (13) imply for $\tau \in \mathbb{F}_0$ that

$$C(x,\tau) = \begin{cases} c_{scr}x + (1-q)\int_0^{\tau} e^{-\delta u}\lambda(u)[c_{se} - c_{scr} - c_{ap}(u)]\mathbb{P}(N_0(u) < x)du \\ +\int_0^{\tau} e^{-\delta u}\lambda(u)[q(c_{se} + c_{re} - c_a(u)) + (1-q)p(u)]du \\ +(h - \delta c_{scr})\int_0^{\tau} e^{-\delta u}\mathbb{E}((x - N_0(u))^+)du + \int_0^{T} e^{-\delta u}\lambda(u)c_a(u)du. \end{cases} \tag{26}$$

By the same argument, we can write relation (14) as

$$C(x,\tau\wedge\sigma_x) = \begin{cases} c_{scr}x + \int_0^{\tau} e^{-\delta u}\lambda(u)[c_{se} + qc_{re} - (1-q)c_{scr} - c_a(u)]\mathbb{P}(N_0(u) < x)du \\ +(h - \delta c_{scr})\int_0^{\tau} e^{-\delta u}\mathbb{E}((x - N_0(u))^+)du + \int_0^{T} e^{-\delta u}\lambda(u)c_a(u)du. \end{cases} \tag{27}$$

To analyze the behavior of both objective functions for a given $\tau \in \mathscr{F}_0$ we need the concept of *discrete convexity*: A function $f : \mathbb{Z}_+ \to \mathbb{R}$ is called *discrete convex* on $\mathbb{Z}_+$ if its first order difference $\Delta_x f(x) := f(x+1) - f(x), x \in \mathbb{Z}_+$ is a non-decreasing function on $\mathbb{Z}_+$. The function f is called *discrete concave* if the function $-f$ is discrete convex. Discrete convexity is applied to our problem through the following result.

Lemma 7. *Let N be a non-homogeneous Poisson process with a locally bounded Borel measurable arrival rate function μ and $f : [0,T] \to \mathbb{R}$ some Borel measurable function. If the function f is non-decreasing and non-positive on $[0,\tau)$ for a given $\tau \in \mathbb{F}_0, 0 < \tau \leq T$ then the function*

$$x \mapsto G(x) := \mathbb{E}\left(\int_0^{\tau\wedge\sigma_x} f(v)\mu(u)du\right) \tag{28}$$

is non-increasing and discrete convex on $\mathbb{Z}_+$. If the function f is non-increasing and non-negative on $[0,\tau)$, then this function is non-decreasing and discrete concave on $\mathbb{Z}_+$.

Proof. See Appendix. □

Using Lemma 7 the next result is easy to verify. A similar result was proved in Frenk et al. (2019b) using a much more complicated proof.

Lemma 8. *Let $\tau \in \mathbb{F}_0, 0 < \tau \leq T$ be given.*

1. *If the procurement cost function c is discrete convex on $\mathbb{Z}_+$ and $c_{ap}(t) \geq c_{se} - c_{scr}$ for every $t < \tau$, then the function $x \mapsto c(x) + C(x,\tau)$ is discrete convex on $\mathbb{Z}_+$.*

2. *If the procurement cost function c is discrete convex on $\mathbb{Z}_+$ and $c_a(t) \geq c_{se} + qc_{re} - (1-q)c_{scr}$ for every $t < \tau$ then the function $x \mapsto c(x) + C(x,\tau\wedge\sigma_x)$ is discrete convex on $\mathbb{Z}_+$.*

Proof. See Appendix. □

Since we always assume (unless stated otherwise) that $p(u) \geq c_{se}$ and c_{ap} is non-increasing, it follows for $c_{scr} \geq 0$ that $c_{ap}(t) \geq c_{se} - c_{scr}$ for every $0 < t \leq T$. Hence by Lemma 8 the function $x \to C(x,\tau)$ is discrete convex for every $0 < \tau \leq T$ and this result was used in Frenk et al. (2019b) to give an efficient algorithm to compute approximately the optimal deterministic policy. In the next section we will only consider the class of pseudo-deterministic policies and prove that one can identify beforehand a finite set of τ values containing an optimal solution for piece-wise constant arrival rate and (decreasing) piece-wise constant substitution cost functions. For this realistic setting, we will analyze the properties of the following optimization problem

$$\upsilon(Q) = \inf_{x \in \mathbb{Z}_+, \tau \in \mathbb{F}_0, 0 \leq \tau \leq T}\{c(x) + C(x, \tau \wedge \sigma_x)\}. \tag{29}$$

3 On properties of the class of pseudo-deterministic policies

In this section we show for a Poissonian defective items stochastic arrival process having a piece-wise constant arrival rate function and a piecewise-constant non-increasing substitution cost function c_a that the finite set of breaking points of these functions contain the optimal switching time between the two used policies and so this optimal switching time can be easily computed. Without loss of generality, we may assume that both functions have the same set of breaking points. If this is not the case, we use the union of both sets of breaking points.

To justify the use of a piece-wise constant substitution cost function we first observe one can always find a piece-wise constant function approximating a bounded non-increasing substitution cost function c_a within any given accuracy using the following procedure. Introducing the composite function $c_{a,n}(t) := d_n(c_a(t)), 0 \leq t \leq T, n \in \mathbb{N}$ with the so-called *dyadic* function $d_n : [0,\infty] \to [0,\infty)$ (cf.Çınlar (2011)) given by

$$d_n(r) = \sum_{k=1}^{n2^n} \frac{k-1}{2^n} 1_{[(k-1)2^{-n}, k2^{-n})}(r) + n 1_{[n,\infty)}(r), \tag{30}$$

yield a sequence of piece-wise constant functions taking only finitely many values. Using this representation we obtain that the function $c_{a,n}$ is again a non-increasing right continuous function satisfying $c_{a,n} \leq c_a$ and for every $n \in \mathbb{N}$, $n > c_a(0)$ we have the approximation errors $\| c_{a,n} - c_a \|_\infty \leq 2^{-n}$ with $\| f \|_\infty = \sup_{t \in [0,T]} | f(t) |$. Similarly, one can also approximate any locally bounded, Borel measurable arrival rate function λ by the function $\lambda_n(t) := d_n(\lambda(t))$ for $t \in [0,T]$ and $\| \lambda_n - \lambda \|_\infty \leq 2^{-n}$ for sufficiently large n. By increasing the value of n, we obtain more accurate approximations. Although not proved in this paper it is relatively easy to give an upper bound on the error in the optimal discounted cost replacing the function c_a by its approximation $c_{a,n}$ and the arrival rate function λ by λ_n. Hence up to any given accuracy one can replace the true functions c_a and λ by their piece-wise constant approximations $c_{a,n}$ and λ_n for some properly selected $n \in \mathbb{N}$. Hence this justifies the use of these piece-wise constant functions.

To solve the optimization problem (29) listed in subsection 2.6 for the set of pseudo-deterministic policies two alternative approaches are possible: The first approach we may use (used in Frenk et al. (2019b) for only the set of static policies) is given by

$$\upsilon(Q) = \inf_{\tau \in \mathbb{F}_0, 0 \leq \tau \leq T} \varphi(\tau),$$

with

$$\varphi(\tau) := \inf_{x \in \mathbb{Z}_+}\{c(x) + C(x, \tau \wedge \sigma_x)\}. \tag{31}$$

Applying this bilevel approach and Lemma 8 we may use the necessary and sufficient first order conditions to determine first the optimal order quantity $x(\tau)$ for each given deterministic switching time $\tau \in \mathbb{F}_0$ and compute for this optimal order quantity the optimal value $\varphi(\tau)$ of optimization problem (31). In general the optimal value function $\varphi : [0,T] \to \mathbb{R}$ is not a convex function on $[0,T]$ and so we cannot use a classical one-dimensional optimization algorithm to determine an optimal $\tau \in \mathbb{F}_0$. However, it can be shown that the function φ is Lipschitz continuous with a computable Lipschitz constant and so up to any given accuracy we may construct a finite discretization $\mathscr{D}$ of the set $[0,T]$. Evaluating the finite number of function values $\varphi(\tau), \tau \in \mathscr{D}$ and taking the minimum of those values we can then approximate $\upsilon(Q)$ up to any given accuracy. Although applicable this approximation approach will not be followed in this paper. The second bilevel approach (considered in this paper) yielding an exact procedure to compute the optimal pseudo-deterministic policy for a piecewise-constant arrival rate and piecewise constant substitution cost function is given by

$$\upsilon(Q) = \inf_{x \in \mathbb{Z}_+} \{\psi(x)\}. \tag{32}$$

with

$$\psi(x) := \inf_{0 \leq \tau \leq T, \tau \in \mathbb{F}_0} \{c(x) + C(x, \tau \wedge \sigma_x)\}. \tag{33}$$

For the optimization problem (32) we first identify the following easy instance.

Lemma 9. *If the substitution cost function c_a of applying the alternative policy is non-increasing and $c_a(0) \leq c_{se} + qc_{re} - (1-q)c_{scr}$ it is optimal within the class of pseudo-deterministic policies not to order any spare parts and switch at time 0 immediately to the substitution policy. This yields the optimal objective value $\upsilon(Q) = \int_0^T e^{-\delta u} \lambda(u) c_a(u) du$.*

Proof. See Appendix. □

Due to the low cost of substitution at time 0 and using relation (14) in the proof of Lemma 9 it is easy to see that it is also optimal to switch immediately at time 0 to the alternative policy within the class of all $(x, \tau \wedge \sigma_x)$ policies with τ an arbitrary bounded stopping time. Hence in the remainder of this paper it is assumed that $c_a(0) > c_{se} + qc_{re} - (1-q)c_{scr}$. In practice this condition on the costs parameters is realistic since immediately applying the alternative policy at time 0 is in general much more expensive then applying the repair-replacement policy. We will now analyze the optimization problem (33) for a Poissonian defective items arrival process having a piece-wise constant arrival rate function and decreasing piece-wise constant substitution or alternative policy cost function $c_a(.)$. To introduce this piece-wise constant arrival rate function consider a strictly increasing sequence $(a_i)_{i=1}^{n+1}$ satisfying $0 = a_1 < a_2 < ... < a_n < a_{n+1} = T$ and define for $t \geq 0$

$$\lambda(t) = \sum_{i=1}^{n} \lambda_i 1_{[a_i, a_{i+1})}(t) \tag{34}$$

for any arbitrarily selected non-negative sequence $(\lambda_i)_{i=1}^{n}$. At the same time the non-increasing substitution cost function of applying the alternative policy is given by

$$c_a(t) = \sum_{i=1}^{n} c_i 1_{[a_i, a_{i+1})}(t), 0 \leq t \leq T. \tag{35}$$

with $c_1 \geq c_2 \geq ... \geq c_n > 0$ and $c_1 > c_{se} + qc_{re} - (1-q)c_{scr}$. Since by Lemma 9 it is optimal not to order and immediately switch to the substitution policy if $c_1 = c_a(0) \leq c_{se} + qc_{re} - (1-q)c_{scr}$, we may assume $c_1 > c_{se} + qc_{re} - (1-q)c_{scr}$. By relation (14) it also follows that the partial derivative of the function $c(x) + C(x, \tau \wedge \sigma_x)$ with respect to $\tau \in \mathbb{F}_0$ for any $a_i < \tau < a_{i+1}, i = 1, ..., n$ is given by

$$\frac{\partial C}{\partial \tau}(x, \tau \wedge \sigma_x) = e^{-\delta \tau} \left(\lambda_i [c_{se} + qc_{re} - (1-q)c_{scr} - c_i] \mathbb{P}(N_0(\tau) < x) + (h - \delta c_{scr}) \mathbb{E}((x - N_0(\tau))^+)\right). \tag{36}$$

Using relation (36) and introducing the index

$$n^* = \max\{1 \leq i \leq n : c_i > c_{se} + qc_{re} - (1-q)c_{scr}\} \leq n, \tag{37}$$

the next result is easy to show for optimization problem (33).

Lemma 10. *Introducing* $\psi_i(x) = \min_{a_i \leq \tau \leq a_{i+1}}\{c(x) + C(x, \tau \wedge \sigma_x)\}$, $i = 1, \ldots, n$ *it follows for every* $x \in \mathbb{Z}_+$ *that* $\psi(x) = \min_{i=1,\ldots,n^*} \psi_i(x)$ *with* $\psi(x)$ *the optimal objective value of optimization problem (33).*

Proof. See Appendix. □

The partial derivative in relation (36) consists of the difference of two positive decreasing functions and so this difference might not be increasing or decreasing. To analyze in detail this partial derivative we introduce the well known Erlang-B formula represented by the function $B : \mathbb{Z}_+ \times \mathbb{R}_+ \to \mathbb{R}$ given by

$$B(x,t) := \frac{\frac{t^x}{x!}}{\sum_{j=0}^{x} \frac{t^j}{j!}}, \; x \in \mathbb{N}, \tag{38}$$

and $B(0,t) = 1$ for every $t > 0$. The value $B(x,t)$ represents the probability that an arriving customer is rejected in a $M/M/x/x$ Markovian loss model with an arrival rate t and departure rate 1 (Kleinrock (1975)). Hence $1 - B(x,t)$ represents the probability that an arriving customer is admitted to the system. By a straightforward computation it is easy to show the following result.

Lemma 11. *If* $N = \{N(t) : t \geq 0\}$ *is a non-homogeneous Poisson process with arrival rate function* θ *and mean arrival function* $\Theta(t) = \int_0^t \theta(u)du, t \geq 0$ *then for any* $x \in \mathbb{N}$ *and* $t > 0$

$$\frac{\mathbb{E}((x - N(t))^+)}{\mathbb{P}(N(t) < x)} = x - \Theta(t)(1 - B(x-1, \Theta(t))). \tag{39}$$

Proof. see Appendix. □

Introducing the function $\kappa_i : \mathbb{Z}_+ \times [0,T] \to \mathbb{R}$ given by

$$\kappa_i(x,\tau) = \lambda_i[c_{se} + qc_{re} - (1-q)c_{scr} - c_i] + (h - \delta c_{scr})x - (h - \delta c_{scr})\Lambda_0(\tau)[1 - B(x-1, \Lambda_0(\tau))], \tag{40}$$

with $\Lambda_0(\tau) = (1-q)\int_0^\tau \lambda(u)du$ and considering relations (36) and (39), we obtain

$$\frac{\partial C}{\partial \tau}(x, \tau \wedge \sigma_x) = e^{-\delta\tau}\mathbb{P}(N_0(\tau) < x)\kappa_i(x,\tau), \tag{41}$$

for $a_i < \tau < a_{i+1}, i = 1, \ldots, n$. For every $x \in \mathbb{N}$ the following result is well known. We list a short proof of this result in the Appendix.

Lemma 12. *For every* $x \in \mathbb{N}$ *the function* $t \mapsto t(1 - B(x,t))$ *is increasing and satisfies* $lim_{t\uparrow\infty} t(1 - B(x,t)) = x$.

Proof. see Appendix. □

Using Lemma 12 we can slightly improve the result in Lemma 10. By Lemma 12 it follows for every $x \in \mathbb{N}$ and $\tau \geq 0$ that $\Lambda_0(\tau)(1 - B(x-1, \Lambda_0(\tau)) \leq x - 1$, and this yields for every $x \in \mathbb{N}$ that the function κ_i listed in relation (40) satisfies

$$\kappa_i(x,\tau) \geq \lambda_i(c_{se} + qc_{re} - (1-q)c_{scr} - c_i) + (h - \delta c_{scr}). \tag{42}$$

Hence it follows for $c_i \leq c_{se} + qc_{re} - (1-q)c_{scr} + \lambda_i^{-1}(h - \delta c_{scr})$ that by relation (41)

$$\frac{\partial C}{\partial \tau}(x, \tau \wedge \sigma_x) = e^{-\delta\tau}\mathbb{P}(N_0(\tau) \leq x)k_i(x,\tau) \geq 0, \tag{43}$$

and this shows the improved result that the function $\tau \mapsto \; c(x) + C(x, \tau \wedge \sigma_x)$ is actually increasing for every x on $[a_i, a_{i+1}]$. This shows for

$$c_1 \leq \; c_{se} + qc_{re} - (1-q)c_{scr} + \lambda_{\max}^{-1}(h - \delta c_{scr}), \tag{44}$$

with $\lambda_{\max} := \max_{i=1,\ldots,n} \lambda_i$ that again it is optimal not to order in optimization problem (32). Hence to solve problem (32) it is sufficient to consider problem instances satisfying

$$c_1 > c_{se} + qc_{re} - (1-q)c_{scr} + \lambda_{\max}^{-1}(h - \delta c_{scr}). \tag{45}$$

For such instances the set $\{1 \leq i \leq n : c_i > c_{se} + qc_{re} - (1-q)c_{scr} + \lambda_{\max}^{-1}(h - \delta c_{scr})\}$ is nonempty and introducing

$$n_0^* := \max\{1 \leq i \leq n : c_i > c_{se} + qc_{re} - (1-q)c_{scr} + \lambda_{max}^{-1}(h - \delta c_{scr})\} \leq n^*, \tag{46}$$

where n^* is listed in (37). Using similar arguments as in Lemma 10, we obtain

$$\psi(x) = \min_{i=1,\ldots,n_0^*} \psi_i(x). \tag{47}$$

In the next result we show for every $x \in \mathbb{Z}_+$ that the finite set of breaking points $a_1, \ldots, a_{n_0^*+1}$ contains an optimal exit time τ of the optimization problem (33).

Lemma 13. *For every $x \in \mathbb{N}$ an optimal solution of optimization problem (33) is attained at some a_i, $i = 1, \ldots, n_0^* + 1$ with n_0^* listed in (46).*

Proof. See Appendix. □

The result in Lemma 13 implies that for any given $x \in \mathbb{Z}_+$ we only have to evaluate the objective function $c(x) + C(x, \tau \wedge \sigma_x)$ at $\tau = a_i$ for $i = 1, \ldots, n_0^* + 1$. Therefore, our optimization problem (32) listed in subsection 2.6 reduces to

$$\upsilon(Q) = \min_{i=1,\ldots,n_0^*+1} \min_{x \in \mathbb{Z}_+} \{c(x) + C(x, a_i \wedge \sigma_x)\}.$$

By Lemma 8 the function $x \to C(x, a_i \wedge \sigma_x)$ is discrete convex for $i \leq n^* + 1$ since $c_a(u) > c_{se} + qc_{re} - (1-q)c_{scr}$ for every $u < a_{n^*+1}$. For a discrete convex procurement function $c(.)$, an optimal solution $x(a_i)$, $i = 1, \ldots, n_0^* + 1$ of $\inf_{x \in \mathbb{Z}_+} \{c(x) + C(x, a_i \wedge \sigma_x)\}$ is given by

$$x(a_i) = \min\{x \in \mathbb{Z}_+ : c(x+1) - c(x) + \Delta C(x, a_i \wedge \sigma_x) \geq 0\}, \tag{48}$$

where $\Delta C(x, a_i \wedge \sigma_x)$ represents the first order operator of the function $x \mapsto C(x, a_i \wedge \sigma_x)$ given by

$$\Delta C(x, a_i \wedge \sigma_x) := C(x+1, a_i \wedge \sigma_{x+1}) - C(x, a_i \wedge \sigma_x), x \in \mathbb{Z}_+. \tag{49}$$

Observe (49) can be used to compute the objective value $c(x(a_i) + C(x(a_i), a_i \wedge \sigma_x)$ using

$$c(x(a_i)) + C(x(a_i), a_i \wedge \sigma_{x(a_i)}) = \sum_{k=0}^{x(a_i)-1} [c(k+1) - c(k) + \Delta C(k, a_i \wedge \sigma_k)] + C(0, a_i \wedge \sigma_0), \tag{50}$$

for every $i = 2, \ldots, n^* + 1$. In (50), we need to evaluate the first order difference operator for different parameter values and the value of the objective function in case of no ordering. Applying relation (27) to the special instances of piece-wise constant arrival rate and substitution cost functions, we obtain the following simplification for the first order cost difference.

Lemma 14. *If the arrival rate function is given by relation (34) and the substitution cost function by relation (35) then for every $i = 1, \ldots, n+1$*

$$C(0, a_i \wedge \sigma_0) = \int_0^T e^{-\delta u} \lambda(u) c_a(u) du = \sum_{j=1}^n \frac{\lambda_j c_j}{\delta} (e^{-\delta a_j} - e^{-\delta a_{j+1}}), \tag{51}$$

and for every $i = 2, \ldots, n+1$ and $x \in \mathbb{Z}_+$

$$\Delta C(x, a_i \wedge \sigma_x) = \begin{cases} \sum_{j=1}^{i-1} \left(\frac{c_{se} + qc_{re} - c_j}{1-q} - c_{scr} \right) [e^{-\delta a_j} \mathbb{P}(N_0(a_j) \leq x) - e^{-\delta a_{j+1}} \mathbb{P}(N_0(a_{j+1}) \leq x)] \\ + (h - \delta c_{scr}) \sum_{j=1}^{i-1} \int_{a_j}^{a_{j+1}} e^{-\delta u} \mathbb{P}(N_0(u) \leq x) du + c_{scr}. \end{cases} \tag{52}$$

Proof. See Appendix. □

For $a_1 = 0$, relation (27) leads to $\Delta C(0,0\wedge\sigma_x) = c_{scr}$. This implies, by relation (48) and $x \to c(x) + \min\{c_{scr},0\}x$ being increasing, that an optimal solution is given by $x(a_1) = 0$ with the optimal value

$$\inf_{x\in\mathbb{Z}_+}\{c(x)+C(x,a_1\wedge\sigma_x)\} = \int_0^T e^{-\delta u}c_a(u)\lambda(u)du = \sum_{j=1}^n \frac{c_j\lambda_j}{\delta}(e^{-\delta a_j} - e^{-\delta a_{j+1}}). \tag{53}$$

To simplify the calculations in Section 5 and determine the optimal pseudo-deterministic policy we will use Lemma 14 together with the following result.

Lemma 15. *If the arrival rate function is given by relation (34) and introducing for every $1 \le j \le n$ and $x \in \mathbb{Z}_+$*

$$\upsilon_j(x) := e^{-\delta a_j}\mathbb{P}(N_0(a_j) \le x),$$

then for every $1 \le j \le n$ and $x \in \mathbb{Z}_+$

$$\int_{a_j}^{a_{j+1}} e^{-\delta u}\mathbb{P}(N_0(u) \le x)du = \frac{\sum_{m=0}^{x}\left(\frac{(1-q)\lambda_j}{\delta+(1-q)\lambda_j}\right)^m[\upsilon_j(x-m) - \upsilon_{j+1}(x-m)]}{\delta+(1-q)\lambda_j}. \tag{54}$$

Proof. See Appendix. □

To solve the optimization problem (32) determining the optimal pseudo-deterministic policy we can now apply the following algorithm.

Solving optimization problem (32) for piece-wise constant arrival rate and substitution cost function with breaking points a_i, $i = 1,\ldots n$ and c discrete convex

- Compute n_0^* in relation (46).
- Solve for every $i = 1,\ldots,n_0^*+1$ the discrete convex minimization problem $\inf_{x\in\mathbb{Z}_+}\{c(x)+C(x,a_i\wedge\sigma_x)\}$ by evaluating the first order conditions $x(a_i) = \min\{x\in\mathbb{Z}_+ : c(x+1)-c(x)+\Delta C(x,a_i\wedge\sigma_x) \ge 0\}$ with $\Delta C(x,a_i\wedge\sigma_x)$ given below

$$\Delta C(x,a_i\wedge\sigma_x) = \begin{cases} c_{scr} + \int_0^{a_i} e^{-\delta u}\lambda(u)[c_{se}+qc_{re}-(1-q)c_{scr}-c_a(u)]\mathbb{P}(N_0(u)=x)du \\ +(h-\delta c_{scr})\int_0^{a_i} e^{-\delta u}\mathbb{P}(N_0(u)\le x)du. \end{cases} \tag{55}$$

- Output

$$\upsilon(Q) = \min_{1\le i\le n^*+1}\{c(x(a_i)) + C(x(a_i),a_i)\},$$

and

$$(x(a_{i^*}),a_{i^*}) = \arg\min_{1\le i\le n^*+1}\{c(x(a_i)) + C(x(a_i),a_i)\}.$$

In Section 5, we present the results of some computational experiments with piece-wise constant arrival rate and substitution cost functions. We determine both the optimal pseudo-deterministic and the optimal stopping time policies for the most general optimization problem (17). Before presenting the computational results, we shortly discuss in the following section how to (approximately) compute the optimal policy of optimization problem (17).

4 On the optimal policy within the class of all stopping times

In this section we discuss how to solve optimization problem (17) listed in Section 2.4. Although most of the results except the characterization of the optimal stopping sets S_k in Lemma 18 and 19 already appeared in Frenk et al. (2019a) using a slightly more difficult approach, we list these main results for completeness also in this paper. By relation (19) we know that there exist a computable upperbound x_U satisfying

$$\upsilon(P) = \inf_{x \leq x_U, x \in \mathbb{Z}_+,} \inf_{\tau \in \mathbb{F}, 0 \leq \tau \leq T} \{c(x) + C(x, \tau)\}. \tag{56}$$

To solve for each $x \leq x_U, x \in \mathbb{Z}_+$ the optimization problem $\varphi(x) := \inf_{\tau \in \mathbb{F}, 0 \leq \tau \leq T} \{c(x) + C(x, \tau)\}$ we discretize the set $[0, T]$ and construct a finite set $\Theta = \{t_0, t_1, ..., t_N\} \subseteq [0, T]$ satisfying $0 = t_0 < t_1 < ... < t_N = T$. Its mesh is given by

$$\Delta(\Theta) = \max_{0 \leq j \leq N-1} \mid t_{j+1} - t_j \mid. \tag{57}$$

Depending on the given set Θ we consider the set of stopping times $\tau \in \mathbb{F}$ taking only values in Θ. For each finite set Θ the set of stopping times only attaining value in Θ is denoted by $\mathbb{F}_\Theta$. Consider now for every $x \leq x_U, x \in \mathbb{Z}_+$ the optimization problem $\varphi_\Theta(x) = \inf_{\tau \in \mathbb{F}_\Theta} \{c(x) + C(x, \tau)\}$, and introduce

$$\upsilon(P_\Theta) = \inf_{x \leq x_U, \tau \in \mathbb{F}_\Theta} \varphi_\Theta(x) = \inf_{x \leq x_U, \tau \in \mathbb{F}_\Theta} \{c(x) + C(x, \tau)\}. \tag{58}$$

Since $\mathbb{F}_\Theta \subseteq \mathbb{F}$ it follows by relation (56) that $0 \leq \upsilon(P_\Theta) - \upsilon(P)$. It is also possibe to construct an upperbound on $\upsilon(P_\Theta) - \upsilon(P)$. If we introduce the functions $f_i : [0, T] \to \mathbb{R}, i = 1, 2$ as follows.

$$f_1(t) = (1 - q)(c_{se} - c_{scr} - c_{ap}(t)),$$

and

$$f_2(t) = q(c_{se} + c_{re} - c_a(t)) + (1 - q)p(t).$$

By using these definitions, one can show the following result. Although this result is shown in Frenk et al. (2019a) we list for completeness an outline of the proof.

Lemma 16. *If the arrival process N of defective items is given by a non-homogeneous Poisson process with arrival rate function $\lambda(.)$ then $\upsilon(P_\Theta) - \upsilon(P) \leq f_0(x_U)\Delta(\Theta)$ and $f_0(x) = h - \delta_{csr}x + \parallel \lambda f_1 \parallel_\infty + \parallel \lambda f_2 \parallel_\infty$ with $\parallel g \parallel_\infty := \sup_{0 \leq t \leq T} \mid g(t) \mid$ denotes the well-known supnorm.*

Proof. see Appendix. □

By Lemma 5 we obtain that the relative error of solving the approximative problem (58) instead of optimization problem (17) is given by

$$1 \leq \frac{\upsilon(P_\Theta)}{\upsilon(P)} \leq 1 + \frac{f_0(x_U)\Delta(\Theta)}{\int_0^T e^{-\delta u} \lambda(u) \min\{c_{se} + qc_{re}, c_a(u)\} du}. \tag{59}$$

In our computational section we select $\varepsilon > 0$ and use a finite set Θ with mesh $\Delta(\theta)$ satisfying

$$0 < \Delta(\Theta) \leq \frac{\varepsilon \int_0^T e^{-\delta u} \lambda(u) \min\{c_{se} + qc_{re}, c_a(u)\} du}{f_0(x_U)}. \tag{60}$$

If (60) holds, then (59) implies that the relative error is smaller than ε. In Section 5, we will use a piece-wise constant arrival rate and a piece-wise constant substitution cost function with the same set of breaking points. Θ is chosen to be $\Theta = \{j\Delta(\Theta) : j = 0, ..., N\}$ with $N\Delta(\Theta) = T$ and the breaking points of arrival rate substitution cost are included in this set. Δ=$\Delta(\Theta) > 0$ we solve the optimization problem (58) using shifted stochastic processes.

Definition 1. *For any non-homogeneous Poisson process X with arrival rate function θ, the shifted stochastic process $\{X^{(k)}(t) : t \geq 0\}$ is defined as $X^{(k)}(t) := X(t+k\Delta) - X(k\Delta)$ for $k = 0, ..., N-1$.*

It is well known that $X^{(k)}$ is again a non-homogeneous Poisson process with arrival rate function $t \to \theta(t+k\Delta)$. Let us define the shifted stochastic process $N_0^{(k)}, k = 0, ..., N-1$ for the arrival process N_0 of non-repairable items with arrival rate function $\lambda_0 = (1-q)\lambda$. For this shifted process the stopping times is defined as $\sigma_x^{(k)} := \inf\{t \geq 0 : N_0^{(k)}(t) \geq x\}$ for every $x \in \mathbb{Z}_+$. At time $k\Delta, k = 0, ..., N-1$ we either stop or continue with the repair-replacement policy. If we decide to stop at time $k\Delta$ and the inventory level equals x, then the random discounted costs of switching to the substitution policy (discounted from time $k\Delta$) is given by $c_{scr}x + \int_0^{(N-k)\Delta} e^{-\delta u} c_a(u+k\Delta) dN^{(k)}(u)$, where $N^{(k)}$ is a shifted Poisson process of arrival requests with rate λ. Hence the expected discounted cost of taking action π_0, which is defined as to switch to the repair-with-substitute policy, at time $k\Delta$ with inventory level x equals

$$C_{\pi_0}(x) := c_{scr}x + \int_0^{(N-k)\Delta} e^{-\delta u} c_a(u+k\Delta)\lambda(u+k\Delta)du. \tag{61}$$

Also the random discounted cost of continuing with the repair-replacement policy at time $k\Delta$ equals $b_k(x) + e^{-\delta\Delta} V_{k+1}((x - N_0^{(k)}(\Delta))^+)$ with $b_k(x)$ the random discounted cost of applying the repair-replacement policy during the time interval $[k\Delta, (k+1)\Delta)$. Introducing

$$B_k(x) = \mathbb{E}(b_k(x)), \tag{62}$$

the expected discounted cost of taking action π_1, defined as not to switch to the repair-with-substitute policy, with inventory level x at time $k\Delta$ becomes

$$C_{\pi_1}(x) := B_k(x) + e^{-\delta\Delta}\mathbb{E}(V_{k+1}((x - N_0^{(k)}(\Delta))^+). \tag{63}$$

This shows for $x \leq x_U$ that the Bellman equations for the above stopping problem are given by the functions $V_k : \mathbb{Z}_+ \to \mathbb{R}$, $k = 0, ..., N$ given by

$$V_N(x) = c_{scr}x, V_k(x) = \min\{C_{\pi_0}(x), C_{\pi_1}(x)\}. \tag{64}$$

To simplify the recurrent relation for V_k in relation (64), we introduce the function $W_k : \mathbb{Z}_+ \to \mathbb{R}$, $k = 0, ..., N$ given by

$$W_k(x) := V_k(x) - \int_0^{(N-k)\Delta} e^{-\delta u} c_a(u+k\Delta)\lambda(u+k\Delta)d. \tag{65}$$

Since it is easy to check that

$$\int_0^{(N-k)\Delta} e^{-\delta u} c_a(u+k\Delta)\lambda(u+k\Delta)du =$$
$$\int_0^{\Delta} e^{-\delta u} c_a(u+k\Delta)\lambda(u+k\Delta)du + e^{-\delta\Delta}\int_0^{(N-(k+1))\Delta} e^{-\delta u} c_a(u+(k+1)\Delta)\lambda(u+(k+1)\Delta)du,$$

we obtain from relations (61)-(65) that we need to solve for $x \leq x_U$ and $k = 0, ..., N$ the Bellman equations

$$W_N(x) = c_{scr}x, W_k(x) = \min\{c_{scr}x, \overline{B}_k(x) + e^{-\delta\Delta}\mathbb{E}(W_{k+1}((x - N_0^{(k)}(\Delta))^+)\}, \tag{66}$$

where

$$\overline{B}_k(x) := B_k(x) - \int_0^{\Delta} e^{-\delta u} c_a(u+k\Delta)\lambda(u+k\Delta)du. \tag{67}$$

For $x=0$ and $k=0,\ldots,N-1$ relation (66) reduces to the simplified recurrent relation

$$W_N(0)=0, W_k(0)=\min\{0,\overline{B}_k(0)+e^{-\delta\Delta}W_{k+1}(0)\}. \tag{68}$$

Observe that after hitting the zero inventory level, so $x=0$, it might be cheaper to continue with the repair-replacement policy depending on the sign of $\overline{B}_k(0)$. This holds if the expected cost $q(c_{se}+c_{re})+(1-q)c_{ap}(u)$ of any defective item arriving at u is much lower then the expected costs $c_a(u)$ of substitution at u. Having computed $W_0(x)$ for $x=0,\ldots,x_U$ with the recurrent relations (66) and (67) after N iterations, we set

$$V_0(x)=W_0(x)+\int_0^{N\Delta}e^{-\delta u}c_a(u)\lambda(u)du, \tag{69}$$

for $x=0,\ldots,x_U$ and solve the optimization problem

$$\min_{x\leq x_U}\{c(x)+V_0(x)\}, \tag{70}$$

by enumeration. To compute $\overline{B}_k(x)$ in (66) we proceed as follows. By definition let $C_k(x,\Delta)$ denote the expected discounted operational cost (discounted from time $k\Delta$) from time $k\Delta$ up to time $N\Delta=T$. If we observe inventory level x at time $k\Delta$ and we apply the deterministic (x,Δ) policy at time $k\Delta$, the shifted arrival processes $N_0^{(k)}$ and $N^{(k)}$ are non-homogeneous Poisson processes with arrival rate functions $t\to\lambda_0(k\Delta+t)$ and $t\to\lambda(k\Delta+t)$. Therefore by the definition of $B_k(x)$ in (62), we need to subtract the discounted substitution costs from time $(k+1)\Delta$ up to $N\Delta=T$ from $C_k(x,\Delta)$ and set the scrapping value equal to zero in the expression for $C_k(x,\Delta)$. This means that

$$B_k(x)=C_k(x,\Delta)-\int_{\Delta}^{(N-k)\Delta}e^{-\delta u}\lambda(k\Delta+u)c_a(u+k\Delta)du, \tag{71}$$

and by relation (26) substituting $c_{scr}=0$

$$C_k(x,\Delta)=\begin{cases}(1-q)\int_0^{\Delta}e^{-\delta u}\lambda(k\Delta+u)[c_{se}-c_{ap}(u+k\Delta)]\mathbb{P}(N_0^{(k)}(u)<x)du\\ +\int_0^{\Delta}e^{-\delta u}\lambda(k\Delta+u)[q(c_{se}+c_{re}-c_a(u+k\Delta))+(1-q)p(k\Delta+u)]du\\ +h\int_0^{\Delta}e^{-\delta u}\mathbb{E}((x-N_0^{(k)}(u))^+)+\int_0^{(N-k)\Delta}e^{-\delta u}\lambda(k\Delta+u)c_a(k\Delta+u)du.\end{cases}$$

This implies using relation (71)

$$B_k(x)=\begin{cases}(1-q)\int_0^{\Delta}e^{-\delta u}\lambda(k\Delta+u)[c_{se}-c_{ap}(u+k\Delta)]\mathbb{P}(N_0^{(k)}(u)<x)du\\ +\int_0^{\Delta}e^{-\delta u}\lambda(k\Delta+u)[q(c_{se}+c_{re}-c_a(u+k\Delta))+(1-q)p(k\Delta+u)]du\\ +h\int_0^{\Delta}e^{-\delta u}\mathbb{E}((x-N_0^{(k)}(u)^+)+\int_0^{\Delta}e^{-\delta u}\lambda(k\Delta+u)c_a(k\Delta+u)du.\end{cases}$$

Applying relation (67) we obtain

$$\overline{B}_k(x)=\begin{cases}(1-q)\int_0^{\Delta}e^{-\delta u}\lambda(k\Delta+u)[c_{se}-c_{ap}(u+k\Delta)]\mathbb{P}(N_0^{(k)}(u)<x)du\\ +\int_0^{\Delta}e^{-\delta u}\lambda(k\Delta+u)[q(c_{se}+c_{re}-c_a(u+k\Delta))+(1-q)p(k\Delta+u)]du\\ +h\int_0^{\Delta}e^{-\delta u}\mathbb{E}((x-N_0^{(k)}(u))^+)du.\end{cases} \tag{72}$$

Hence by relation (72) and the definition of the first order difference operator, the next result easily follows.

Lemma 17. *For every $k \in \mathbb{N}$ it follows that*

$$\overline{B}_k(0) = \int_0^{\Delta} e^{-\delta u}\lambda(k\Delta+u)[q(c_{se}+c_{re}-c_a(k\Delta+u))+(1-q)p(k\Delta+u)]du, \tag{73}$$

and for every $x \in \mathbb{Z}_+$

$$\Delta\overline{B}_k(x) = (1-q)\int_0^{\Delta} e^{-\delta u}\lambda(k\Delta+u)[c_{se}-c_{ap}(k\Delta+u)]\mathbb{P}(N_0^{(k)}(u)=x)du + h\int_0^{\Delta} e^{-\delta u}\mathbb{P}(N_0^{(k)}(u)\leq x)du. \tag{74}$$

By relation (66) we obtain the following optimal stopping sets S_k.

$$S_k := \{x \in \mathbb{Z}_+ : c_{scr}x \leq \overline{B}_k(x) + e^{-\delta\Delta}\mathbb{E}(W_{k+1}((x-N_0^{(k)}(\Delta))^+))\}, \tag{75}$$

for $k = 0,\ldots,N-1$. Under certain intuitively clear conditions on the arrival rate and the substitution cost one can show a desirable property for the optimal stopping set S_k. To this end, we first need the next two results.

Lemma 18. *If the arrival process of defective items is a homogeneous Poisson process and the penalty costs are constant over $[0,T]$ then for every $0 \leq k \leq N-1$ and for every $x \in \mathbb{Z}_+$*

$$W_k(x) \leq W_{k+1}(x) \leq c_{scr}x.$$

Proof. see Appendix. □

The following result follows from Lemma 18.

Lemma 19. *If the arrival process of defective items is a homogeneous Poisson process and the penalty costs are constant over $[0,T]$ then $S_k \subseteq S_{k+1}$ for every $0 \leq k \leq N-1$.*

Proof. see Appendix. □

By Lemma 19 we obtain that for every time $0 \leq k \leq N-1$ there exists no threshold value or there exists some threshold inventory level $x_k^\star$ satisfying we will always stop at time k if and only if our inventory level is above this level. In Section 5, we numerically find that the same structure for our optimal stopping sets in Lemma 19 holds for our piece-wise homogeneous Poisson process. This is probably due to Δ being extremely small, so we have $\mathbb{P}(N_0^{(k)} \leq 1) \simeq 1$ at the breaking points which is partly indicated by the proof of Lemma 18. Hence due to (72) the function $x \mapsto \overline{B}_k(x)$ is almost an affine function with approximately the same slopes depending mainly on the inventory cost h and the discount factor δ but with increasing constant terms.

5 Computational results

In this section we numerically analyze the performance of the optimal stopping problem (17) and determine the optimal pseudo-deterministic policies for different parameter settings. We measure the sensitivity of the performance gap between pseudo-deterministic and deterministic policies to different parameters. Our computational experiments are coded in R Gui and run with a 1.60GHz processor. It took on average 8 minutes to execute the algorithm to compute the optimal stopping policy for the dynamic model and 4 seconds to run the algorithm for the optimal pseudo-deterministic policy.

5.1 Experimental setting and base scenario

In our test bed, the planning horizon $[0,T]$ is split into three equal intervals by the sequence of breaking points $(a_i)_{i=1}^4$, where $a_i=(i-1)\frac{T}{3}$. Therefore, our sub-intervals are $[0,\frac{T}{3}]$, $[\frac{T}{3},\frac{2T}{3}]$ and $[\frac{2T}{3},T]$, and in each of these sub-intervals the arrival rate and the substitution cost functions are assumed to be equal to λ_i and c_i for $i=1,2,3$ respectively. The cost of the repair-with-substitute policy over different sub-intervals is now given by $c_j=c_a(0)e^{-\gamma a_j}$ for $j=1,2,3$, where γ is the decay factor. The arrival rate of the defective products is assumed to be a step function; constant arrival rate of $\lambda_i=l\beta^{i-1}$ over the interval $[a_i,a_{i+1}]$ and jumps at time a_i. For a given β, l is set to a value that makes the expected number of product arrivals over the horizon $[0,T]$ equal to $10T$ (i.e., on average 10 customer requests per unit time). For the general (x,τ)-policy the penalty costs, p, is assumed to be equal to 1290. The base case scenario with these parameters is given in Table 1. Note that our substitution cost and arrival rate functions are also used by Frenk et al. (2019a), to which we benchmark our results.

Table 1: Problem parameters for the base case scenario.
width=0.9

T	c_{scr}	c_{se}	c_{re}	h	q	$\bar{c}$	$c_a(0)$	γ	p	ε	δ	β
66 (months)	30	30	20	3.25	0.5	225	645	0.02	1290	0.001	0.003	0.5

The procurement cost function has the form $c(x)=\bar{c}x$ with rate $\bar{c}>0$. Also we assume $\bar{c}>c_{scr}^-=-\min\{c_{scr},0\}$ which makes $x\mapsto c(x)-c_{scr}^-x$ an increasing non-negative function with limit ∞ satisfying our standard assumption on the cost parameters. This also implies that the first order conditions in (48) reduce to

$$x(a_i)=\min\{x\in\mathbb{Z}_+:\Delta C(x,a_i\wedge\sigma_x)\geq-\bar{c}\}. \tag{76}$$

To solve the optimization problem (17), we apply the approximation approach proposed in Section 4 and select the finite set Θ having mesh $\Delta>0$ in such a way that the breaking points are contained in Θ. This means that $M_i\Delta=a_i, i=1,...,n+1$ and so $M_1=0$ and $M_{n+1}=N$. Selecting a relative error of $\varepsilon=0.001$ in the base case scenario, the mesh Δ of our set Θ is not larger than 0.003 due to (60). Due to the approximation approach, the computed optimal objective values $\upsilon(P)$, in Table 2, are subject to this selected relative error. For the above settings one can also simplify the computation of $\Delta\overline{B}_k(x)$. The next result follows immediately from Lemma 17.

Lemma 20. *If the arrival rate and substitution policy cost functions are given by relations (34), and (35) then for $M_i\leq k\leq M_{i+1}-1, i=1,...,n+1$*

$$\overline{B}_k(0)=[q(c_{se}+c_{re}-c_i)+(1-q)p]\lambda_i\delta^{-1}(1-e^{-\delta\Delta}). \tag{77}$$

Also for every $m\in\mathbb{Z}_+$

$$\Delta\overline{B}_k(m)=[h+\delta(p+c_i-c_{se})]\int_0^\Delta e^{-\delta u}\mathbb{P}(X(u)\leq m)du-(p+c_i-c_{se})[1-e^{-\delta\Delta}\mathbb{P}(X(\Delta)\leq m)]. \tag{78}$$

and X a homogeneous Poisson process with arrival rate $(1-q)\lambda_i$.

Solving the Bellman equations in relations (66) and (67) we are able to compute the function V_0 in relation (69) and solve optimization problem $\min_{x\leq x_U}\{c(x)+V_0(x)\}$. In Figure 2 the function V_0 is plotted for the values $x=0,1,...,x_U$ under the base case scenario. This plot shows that this function achieves a minimum within this range.

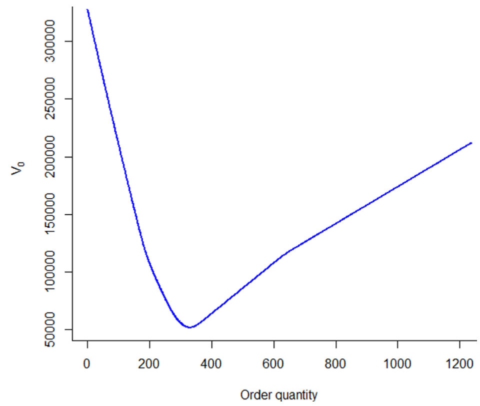

Figure 2: Function $V_0(x)$ plotted over the range of x values for the base case scenario.

5.2 Sensitivity of $\upsilon(P)$ to problem parameters

The optimal solution of (17) consists of the optimal initial order quantity, x, and the optimal stopping time to switch from the repair-replacement policy to the repair-with-substitute policy, τ. Plainly, the set of points (x, τ) represent the times τ at which it is optimal to stop holding inventory x. In order to provide insight into the structure of the optimal policy, we present the optimal (x, τ) policies as blue regions in Figure 3 for different values of holding cost and decay factors. If for a given realization of the arrival process one enters the blue region at a certain time with x items in inventory, it is optimal to switch at that time to the repair-with-substitute policy. Also in the base case scenario it follows for every $1 \leq i \leq n$ that $q(c_{se} + c_{re} - c_i) + (1-q)p = 670 - 322.5\theta^{i-1}$. This shows that the conditions of Lemma (1) are satisfied restricting our class of optimal policies for the optimization problem (17). At the optimal stopping time we immediately switch to the substitution policy (from the repair-replacement policy) at the moment the spare parts inventory is zero. This is depicted by the thin blue horizontal lines at the bottom of each plot in Figure 3.

In Figure 3, we observe that as γ increases, the substitution policy cost declines faster and that makes switching to the substitution policy by scrapping inventory more profitable. h has similar but opposite effect on the stopping region. With the increase of h, the repair-replacement policy becomes more expensive. Hence, the size of stopping area is increasing in h. These results are consistent with the results of Frenk et al. (2019a). The main difference between their results and ours stems from the cost of substitution policy, which is larger in our experiments. Therefore, we encounter larger stopping regions than Frenk et al. (2019a) in all scenarios.

In Figure 7, we depict the effect of different parameter values on the optimal cost and order quantity of the problem 17. In each plot, the value of a parameter is changed whereas the rest of the parameters are set to the values in Table 1. The optimal cost value is depicted with a dashed line whereas the optimum order size is presented with a straight line on the secondary axis.

Figure 4 indicates that the optimal order size and the optimum cost decrease in γ with different rates. Decrease in the optimum cost values occurs with a constant rate due to decaying cost parameters. The optimum order size converges to 200 for $\gamma \geq 0.05$ due to the fact that order size of less than 200 lead to high penalty costs.

Figure 5 also depicts the effect of the probability q on the optimum order size and the optimum cost value. As q increases, less spares will be needed and thus the optimum order is smaller. In addition, we apply the repair-replacement policy for most of the time horizon which in turn incurs lower expected total cost. For the extreme case $q = 1$, which represents all items can be repaired, the optimal order quantity is 0 and the total expected cost of the optimal policy is 30.757 thousand. On the other hand, for $q = 0$,

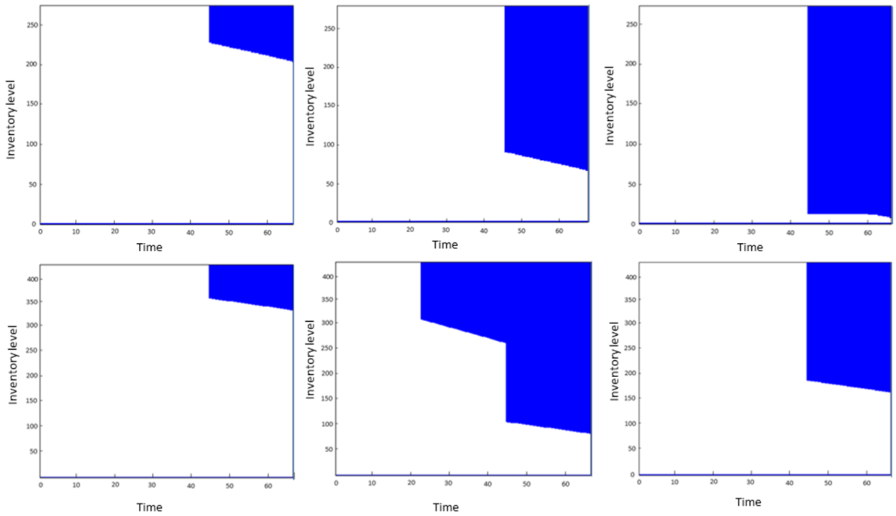

Figure 3: Stopping regions for different γ and h values. First row plots are for $\gamma \in \{0.03, 0.05, 0.07\}$. Second row plots are for $h \in \{3.25, 6.5, 13\}$.

we order initially a large quantity to cover the demand and therefore, we incur a large optimum expected cost. This shows that the effect of the repair probability is rather significant on the system behavior and optimum results. Figure 6 shows the relation between the optimum results and the inventory holding cost rate. Expectedly, increasing values of h lead to lower order size and higher total cost in the optimum solutions.

5.3 Performance of pseudo-deterministic policy for different parameters

To investigate the effect of different problem parameters on $\upsilon(P)$ and $\upsilon(Q)$, we conduct numerical experiments by altering parameter values of the base scenario in one-at-a-time fashion. For each parameter setting, we report the optimal values of $\upsilon(P)$ and $\upsilon(Q)$ together with their relative percent difference given below:

$$RPD(P:Q) = 100 \times \frac{\upsilon(Q) - \upsilon(P)}{\upsilon(P)}. \tag{79}$$

In addition, we report the optimal initial order quantities for both optimization problems and switching times for $\upsilon(Q)$. The results of our numerical experiments are reported in Table 2.

In all considered scenarios the optimal policies of $\upsilon(P)$ and $\upsilon(Q)$ are almost the same and this explains the small relative deviation. The relative error between the optimal objective $\upsilon(P)$ and $\upsilon(Q)$ is on average 0.15 percent. Such a low difference justifies using the optimal pseudo-deterministic policy as an approximation to the general optimization problem. At the same time, the small difference between the two policy shows that one cannot prove that the optimal pseudo-deterministic policy is optimal within the larger class of stopping times and that within the class of pseudo-deterministic policies the order quantity is the most important decision variable to control the cost. For many parameter settings, the optimal order quantity of the problem (17) satisfies

$$x_{opt} \leq \min\{x \in \mathbb{Z}_+ : (x, \tau) : W_\tau(x) = c_{scr}x\}, \tag{80}$$

showing that the optimal policy never switches to the repair-with-substitute policy before the end of the horizon with a positive inventory. The switch to substitution takes place before the end of the horizon

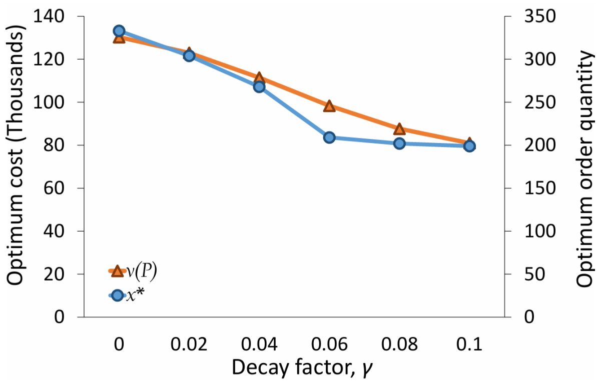

Figure 4: Effect of decay rate (γ)

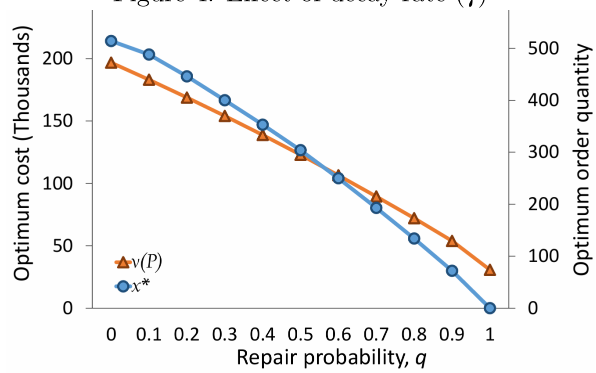

Figure 5: Effect of non-repairable probability (q)

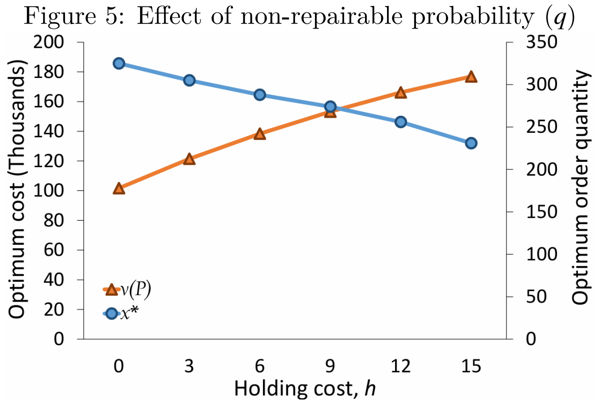

Figure 6: Effect of holding cost (h)

Figure 7: Sensitivity of the optimum solution of the problem 17 to different parameters.

with 0 inventory. This means that the optimal dynamic policy actually coincides with the optimal pseudo-deterministic policy, explaining the extremely small relative deviation, which is mainly caused by our approximation to $\upsilon(P)$. For the pseudo-deterministic policy, we calculate the probability of $\{\tau_0 \leq \sigma_x\}$, which represents the event of switching to the repair-with-substitute policy with positive inventory. Table 2 presents the calculated probabilities, which are very close to zero in almost all cases. For the cases failing to satisfy (80), the highest probabilities of switching with positive inventory are related to high substitution costs either directly by means of $c_a(0)$ or indirectly through γ and q. The probability is almost 0 in case of low h, high initial $c_a(0)$ or high q. All of these findings are intuitively clear. In addition, we computed the probability of switching to substitution with a positive spare parts inventory in the optimum solution of (17). This probability, which we didn't report in Table 2 due to space limitation, is again found to be extremely small. Hence in both optimal policies (recall that we will only stop at the end of the horizon with a positive inventory level under the pseudo-deterministic policy) switching to the repair-with-substitute policy with a positive spare parts inventory occurs rarely when the system starts with the optimal order quantity.

Generally we observe that time to switch to substitution coincides with the breaking point at which substitution becomes cheaper and inventory is depleted in the optimal policies. In Table 2 we observe that the optimal order quantities of both policies are the same and the objective values of the two problems are very close to each other. This is partly due to the fact that the majority of the cases in Table 2 satisfy the condition in Lemma 1. In majority of the cases, the deterministic switching time of the optimal pseudo-deterministic policy is equal to 66, i.e. $\tau_{opt} = 66$ except two cases where the substitution is relatively inexpensive at time 0. In those cases $\tau_{opt} = 44$. $\tau_0 = 66$ represents the utilization of the repair-replacement policy until σ_x, after which substitution is utilized.

In Table 2, we notice that the decay factor, γ, of the substitution cost function has a negative effect on order sizes and costs of both optimal policies. Decreasing values of γ make substitution expensive leading to a higher total expected cost. For the same reason, we order a higher initial inventory so that we can avoid switching to the repair-with-substitute policy due to reaching earlier inventory level 0. The initial substitution cost $c_a(0)$ has a similar effect on the optimal cost and the optimal initial inventory. In our analyses we find that the optimum cost and policy parameters are insensitive to c_{scr} value. This is mainly because switch time mostly corresponds to σ_x in almost all cases. Similarly, p values do not have major effect on the objective value as long as (16) in Lemma 1 is satisfied (cases that do not satisfy (16) are evaluated in Section 5.5).

Furthermore, we find that q has a large negative effect on the expected total costs for both policies. In case of high q, a lot of the returned products need only a repair service, which leads to smaller order size and lower cost. Parameters h and β have similar effect on the optimal order quantity and the objective value. Expectedly, a higher h increases the average cost of the repair-replacement policy and decreases the optimum order size. For higher values of β most of the arrivals happen in later periods (recall that the total number of expected arrival is $10T$). This decreases the optimal order quantity as substitution gets cheaper over time.

5.4 End-of-life management for different product types

We also compute the optimal dynamic policy and the optimal pseudo-deterministic policy for different lengths of the warranty period by considering three different product types. Each of these cases is presented in Table 3 and the different length T refers to a certain type of product. First, we compare the performance of the two policies on the second product type in the table (row with $T = 66$). The box plot in Figure 8 indicates a very small and always positive difference between the optimal policy in problem (17) and the optimal pseudo-deterministic policy. Also one can observe that both policies order the same amount of spare parts at the beginning of the planning horizon.

The first product in the table represents a high-demand electronic device with $T = 24$ months of warranty. The product type can be exemplified with cell phones. Since, the product has a short life cycle,

Table 2: Sensitivity analysis of the problem parameters for different policies.

		$\upsilon(P)$		$\upsilon(Q)$			$\mathbb{P}(\tau_0 < \sigma_x)$	$RPD(P:Q)$
		x	cost	x	τ_0	cost		(%)
γ	0.05	227	105,455.8	227	66	105,552.9	1.19×10^{-9}	0.09
	0.02	304	122,965.6	304	66	122,974.6	7.88×10^{-2}	0.09
	0.01	323	127,140.5	323	66	127,259.6	3.64×10^{-1}	0.09
	0.005	329	128,755.6	329	66	128,878.2	4.93×10^{-1}	0.10
$c_a(0)$	250	191	100,289.7	191	44	100,382.0	4.16×10^{-9}	0.09
	322.5	216	109,735.0	216	44	109,732.0	2.01×10^{-5}	0.09
	1290	329	128,991.9	329	66	129,111.9	4.93×10^{-1}	0.09
	2580	342	132,955.64	342	66	133,083.2	7.56×10^{-1}	0.10
p	322.5	304	122,845.2	304	66	122,974.6	7.88×10^{-2}	0.11
	645	304	122,850.3	304	66	122,974.6	7.88×10^{-2}	0.10
	2580	304	122,854.2	304	66	122,974.6	7.88×10^{-2}	0.10
	5160	304	122,853.5	304	66	122,974.6	7.88×10^{-2}	0.10
h	0.8125	320	107,237.1	320	66	107,350.5	3.03×10^{-1}	0.11
	1.625	315	112,679.2	315	66	112,787.6	2.14×10^{-1}	0.10
	6.5	285	140,809.4	285	66	140,953.0	6.24×10^{-3}	0.10
	13	248	169,876.7	248	66	170,039.8	1.41×10^{-6}	0.10
c_{scr}	−30	304	122,829.7	304	66	122,945.8	7.88×10^{-2}	0.09
	10	304	122,841.2	304	66	122,965.0	7.88×10^{-2}	0.10
	60	304	122,864.2	304	66	122,989.1	7.88×10^{-2}	0.10
	90	303	122,879.7	303	66	123,001.8	7.09×10^{-2}	0.10
q	0.4	353	138,842.7	353	66	138,849.5	1.51×10^{-2}	0.10
	0.6	250	106,489.7	250	66	106,500.5	2.04×10^{-1}	0.09
	0.8	134	72,127.8	134	66	72,145.2	5.92×10^{-1}	0.13
	1	0	30,772.5	1	66	31,232	1	1.54
β	0.25	319	119,943.2	319	66	119,958.8	2.84×10^{-1}	0.10
	1	270	130,079.9	270	66	130,078.7	3.73×10^{-4}	0.09
	1.5	244	134,660.0	244	66	134,652.9	4.20×10^{-7}	0.09
	2	227	137,495.7	227	66	137,485.0	1.19×10^{-9}	0.09

the average number of service requests per unit of time $\Lambda(T)/T$, the substitution cost and its decay factor are expected to be high. The second row in Table 3 represents products with medium economic lifetime (5-10 years) with relatively cheaper substitution and repair costs. As it has longer lifetime, its repair demand is slower than the electronic products. This product type can be exemplified with household appliances such as dishwasher. The last row refers to a product with a long warranty period and very low value of $\Lambda(T)/T$. An example of such a product can be an expensive medical machine/equipment. The average monthly service demand as well as the decay of the substitution cost are very low for these types of products. The box plot of the operational costs is given in Figure 9. The box plot shows that short and long-life products have closer optimum cost whereas the medium-life product has a much lower optimum cost. The proximity of the optimum costs of short and long-life products, despite the difference between the cost rates, mainly stems from the difference between their demand rates. Lower optimum cost of medium-life products, on the other hand, is due to cheaper substitution, service and repair costs.

In order to investigate the relative cost difference between (17) and (32), we calculate the statistics in

Table 3: Three types of products with different characteristics.

Type	T	$\Lambda(T)/T$	parameter ranges
Short life, medium cost (Cell phone)	24	50	$c_{scr} \in [-10,80]$, $c_{se} \in [20,70]$, $c_{re} \in [10,70]$, $h \in [0.5,1.5]$, $q \in [0.2,0.5]$, $c \in [250,750]$, $c_a(0) \in [1250,3000]$, $\gamma \in [0.025,0.05]$, $p \in [1000,3000]$, $\delta \in [0.001,0.01]$, $\beta \in [0.7,1]$
Medium life, low cost (Dishwasher)	66	10	$c_{scr} \in [-10,40]$, $c_{se} \in [10,50]$, $c_{re} \in [10,40]$, $h \in [1,5]$, $q \in [0.3,0.7]$, $\bar{c} \in [100,400]$, $c_a(0) \in [750,2000]$, $\gamma \in [0.01,0.03]$, $p \in [500,3000]$, $\delta \in [0.001,0.01]$, $\beta \in [0.5,0.7]$
Long life, high cost (MRI Scanner)	120	1	$c_{scr} \in [-1000,2000]$, $c_{se} \in [200,700]$, $c_{re} \in [1000,3000]$, $h \in [10,30]$, $q \in [0.5,0.9]$, $c \in [7500,20000]$, $c_a(0) \in [15000,30000]$, $\gamma \in [0.005,0.015]$, $p \in [10000,30000]$, $\delta \in [0.001,0.01]$, $\beta \in [0.5,0.7]$

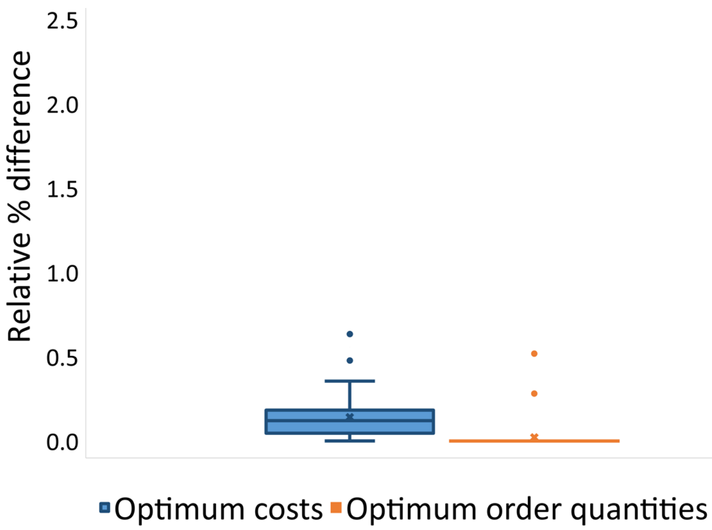

Figure 8: Difference between the dynamic and the pseudo-deterministic policies for a medium-life product ($T = 66$).

(79) for each generated parameter sets for the three product types. Percent cost difference between the two problems are depicted in Figure 10. Results of our calculations indicate that the pseudo-deterministic policy provides a very good approximation to the optimal policy for short-life products. However, as product lifetimes increase (while spare parts demand decreases), the cost difference increases up to 300%. This indicates that the pseudo-deterministic policy has a questionable performance for long-life capital products due to the importance of penalty costs, which is ignored by the pseudo-deterministic policy.

5.5 Worst case performance of pseudo-deterministic policy for short-life products.

In Sections 5.2 and 5.3, we consider parameter settings that mostly satisfy the condition (16) in Lemma 1 and similar to short-life medium-cost products in Table 3. Specifically, 24 out of 28 cases in Table 2 satisfy (16) in Lemma 1, which proves that the pseudo-deterministic policy is optimal for the GLTB problem. In order to investigate the worst case performance of the pseudo-deterministic policy, we conduct numerical experiments with the parameter values, given in Table 4, which violate the condition (16). To this end,

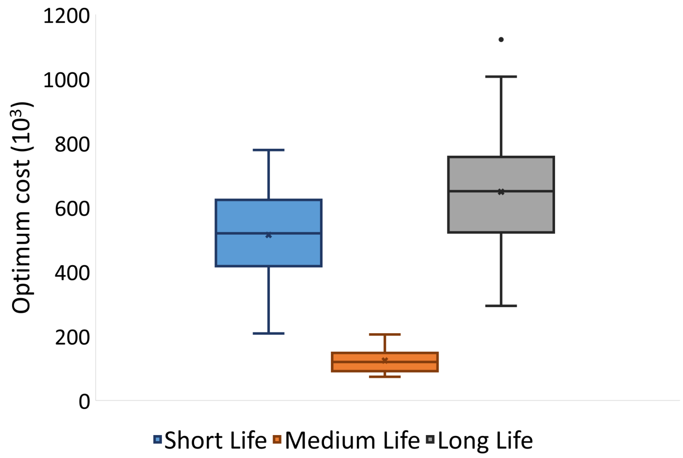

Figure 9: Operational costs for the three different product types in Table 3.

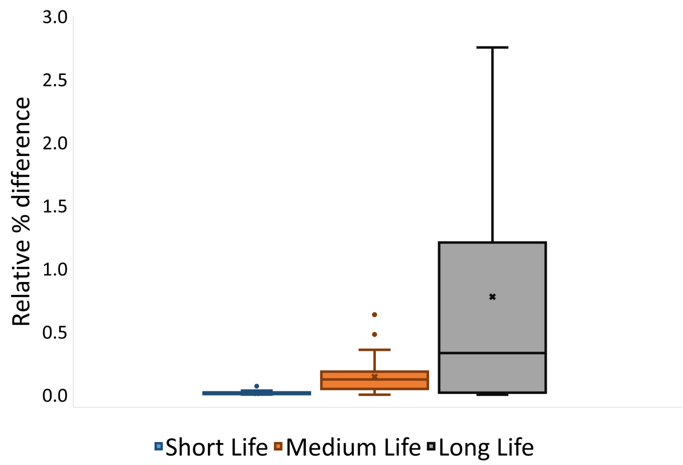

Figure 10: Percentage of relative cost efficiency for the three product types.

we set q to the values close to 1 and choose $c_a(0)$ such that (16) fails at least one part of the planning horizon. Note that all unspecified parameters are equal to the values given in Table 1.

Our results in Table 4 indicate that pseudo-deterministic policy performs near-optimally. The difference between the optimal policy and the pseudo-deterministic policy appears when we $q = 0.999$. Even in those cases the cost difference does not exceed 0.02 %. Recall that the optimal solution of the GLTB problem is presented in Lemma 2 for $q = 1$. Therefore, the parameters of the pseudo-deterministic solution can easily be modified for those cases using our analytical results (Lemma 2).

Table 4: Worst case performance of pseudo-deterministic policy.

q	$c_a(0)$	$x^*(P)$	$x^*(Q)$	$\tau^*(Q)$	$RPD(Q:P)$
0.999	35	0	0	0	0.00
	70	0	1	22	0.01
	105	0	1	44	0.01
	140	0	1	66	0.02
0.95	70	16	16	22	0.00
	120	27	27	44	0.00
	140	29	29	66	0.00
	210	34	34	66	0.00
	280	36	36	66	0.00
0.9	120	44	44	66	0.00
	240	62	62	66	0.00
	360	67	67	66	0.00
	480	70	70	66	0.00

6 Conclusion

This paper is a continuation of Teunter and Fortuin (1999), Pourakbar et al. (2012), Frenk et al. (2019b) and Frenk et al. (2019a). In Teunter and Fortuin (1999) it is assumed that during the so-called end-of-life phase of a product one always uses the so-called repair-replacement policy and never switches to an alternative policy. In many practical problems, other end-of-life management strategies are utilized to avoid (possibly) high holding costs of Last Time Buy orders. Finding a substitute part is recognized to be a good alternative policy that can be executed. As the cost of such a policy becomes cheaper over time Pourakbar et al. (2012) proposed the idea of switching to a alternative policy instead of holding inventory for the rest of the planning horizon. Accordingly, they extended the class of policies and proposed a heuristic procedure to determine a optimal policy among this larger class. Subsequently for the important subclass of deterministic policies Frenk et al. (2019b) gives an exact procedure to determine the optimal deterministic policy. Also, in Frenk et al. (2019a) an algorithm was proposed to identify the optimal so-called dynamic policy by studying in more detail the problem originally considered in Pourakbar et al. (2012). In the current paper we consider the same problem and introduce among the set of all dynamic policies the subclass of pseudo-deterministic policies. Under the assumption that the substitution policy cost function and the arrival intensity function are piece-wise constant functions we show it is easy to identify the optimal pseudo-deterministic policy. In our computational section we then compare the objective value of the optimal pseudo-deterministic policy with the optimal objective value of the more general optimal stopping problem and show numerically for different scenarios that these objective values are close. This can be explained since the class of deterministic polices avoid possibly high penalty costs and in this case an optimal policy avoids these penalty costs. This empirical evidence and its intuitive explanation suggests that the simpler set of pseudo-deterministic policies can serve as an approximation of the optimal policy within the more general optimal stopping problem.This justifies a detailed study of these class of policies.

Acknowledgement

The authors like to thank the anonymous referees for their constructive comments which improved the presentation of the paper.

References

Ahmad, S., and Butt, M. M. (2012). Can after sale service generate brand equity? *Marketing Intelligence & Planning*, **30**, 307–323.

Bayındır, Z. P., Erkip, N., and Güllü, R. (2007). Assessing the benefits of remanufacturing option under one-way substitution and capacity constraint. *Computers and Operations Research*, **34**, 487–514.

Behfard, S., van der Heijden, M. C., Al Hanbali, A., and Zijm, W. H. (2015). Last time buy and repair decisions for spare parts. *European Journal of Operational Research*, **244**, 498–510.

Bradley, J. R., and Guerrero, H. H. (2008). Product design for life-cycle mismatch. *Production and Operations Management*, **17**, 497–512.

Bradley, J. R., and Guerrero, H. H. (2009). Lifetime buy decisions with multiple obsolete parts. *Production and Operations Management*, **18**, 114–126.

Çınlar, E. (2011). *Probability and Stochastic.* Vol 261, Springer Science & Business Media.

Cohen, M. A., Agrawal, N., and Agrawal, V. (2006). Winning in the aftermarket. *Harvard Business Review*, **84**, 129.

Dietrich S. and Daryl J. D. (1983). *Comparison Methods for Queues and Other Stochastic Models.* Wiley, New York.

Fortuin, L. (1980). The all-time requirement of spare parts for service after sales-theoretical analysis and practical results. *International Journal of Operations & Production Management*, **1**, 59–70.

Fortuin, L. (1981). Reduction of all-time requirements for spare parts. *International Journal of Operations & Production Management*, **2**, 29–37.

Frenk, J. B. G., Javadi, S., Pourakbar, M., and Sezer, S. O. (2019). An exact static solution approach for the service parts end-of-life inventory problem. *European Journal of Operational Research*, **272**, 496–504.

Frenk, J. B. G., Javadi, S., Pourakbar, M., and Sezer, S. O. (2019b). An exact static solution approach for the service parts end-of-life inventory problem. *European Journal of Operational Research*, **272**, 496–504.

van der Heijden, M., and Iskandar, B. P. (2013). Last time buy decisions for products sold under warranty. *European Journal of Operational Research*, **224**, 302–312.

Hekimoğlu, M., van der Laan, E., and Dekker, R. (2018). Markov-modulated analysis of a spare parts system with random lead times and disruption risks. *European Journal of Operational Research*, **269**, 909–922.

Huang, W., Kulkarni, V., and Swaminathan, J. M. (2007). Coordinated inventory planning for new and old products under warranty. *Probability in the Engineering and Informational Sciences*, **21**, 261–287.

Huang, W., Kulkarni, V., and Swaminathan, J. M. (2008). Managing the inventory of an item with a replacement warranty. *Management Science*, **54**, 1441–1452.

Hur, M., Keskin, B. B., and Schmidt, C. P. (2018). End-of-life inventory control of aircraft spare parts under performance based logistics. *International Journal of Production Economics*, **204**, 186–203.

Inderfurth, K., and Kleber, R. (2013). An advanced heuristic for multiple-option spare parts procurement after end-of-production. *Production and Operations Management*, **22**, 54–70.

Inderfurth, K., and Mukherjee, K. (2008). Decision support for spare parts acquisition in post product life cycle. *Central European Journal of Operations Research*, **16**, 17–42.

Javadi S. (2018). On the end-of-life inventory problem (Doctoral dissertation). PhD thesis, Sabanci University, Istanbul, Turkey.

Kim, B., and Park, S. (2008). Optimal pricing, EOL (end of life) warranty, and spare parts manufacturing strategy amid product transition. *European Journal of Operational Research*, **188**, 723–745.

Van Kooten, J. P., and Tan, T. (2009). The final order problem for repairable spare parts under condemnation. *Journal of the Operational Research Society*, **60**, 1449–1461.

Kleinrock, L. (1975). *Queueing Sytems.* Vol 1. Wiley, New York.

Krikke, H., and van Der Laan, E. (2011). Last Time Buy and control policies with phase-out returns: a case study in plant control systems. *International Journal of Production Research*, **49**, 5183–5206.

Li, X., Dekker, R., Heij, C., and Hekimoğlu, M. (2016). Assessing end-of-supply risk of spare parts using the proportional hazard model. *Decision Sciences*, **47**, 373–394.

Padmanabhan, V., and Rao, R. C. (1993). Warranty policy and extended service contracts: Theory and an application to automobiles. *Marketing Science*, **12**, 230–247.

Pourakbar, M., van derLaan, E., and Dekker, R. (2014). End-of-life inventory problem with phaseout returns. *Production and Operations Management*, **23**, 1561–1576.

Pourakbar, M., Frenk, J. and Dekker, R, (2012). End-of-life inventory decisions for consumer electronics service parts. *Production and Operations Management*, **21**, 889–906.

Ross, S. M. (1997). *Introduction to Probability Models.* Academic Press, New York.

Sahyouni, K., Savaskan, R. C., and Daskin, M. S. (2010). The effect of lifetime buys on warranty repair operations. *Journal of the Operational Research Society*, **61**, 790–803.

Shen, Y., and Willems, S. P. (2014). Modeling sourcing strategies to mitigate part obsolescence. *European Journal of Operational Research*, **23**, 522–533.

Shi, Z. (2019). Optimal remanufacturing and acquisition decisions in warranty service considering part obsolescence. *Computers and Industrial Engineering*, **135**, 766–779.

Shi, Z., and Liu, S. (2020). Optimal inventory control and design refresh selection in managing part obsolescence. *European Journal of Operational Research*, **287**, 133–144.

Solomon, R., Sandborn, P. A., and Pecht, M. G. (2000). Electronic part life cycle concepts and obsolescence forecasting. *IEEE Transactions on Components and Packaging Technologies*, **23**, 707–717.

Teunter, R. H., and Fortuin, L. (1999). End-of-life service. *International Journal of Production Economics*, **59**, 487–497.

Teunter, R. H., and Haneveld, W. K. K. (1998). The final orders problem. *European Journal of Operational Research*, **107**, 35–44.

Westerweel, B., Basten, R. J., and van Houtum, G. J. (2018). Traditional or additive manufacturing? Assessing component design options through lifecycle cost analysis. *European Journal of Operational Research*, **270**, 570–585.

A Appendix: Proof of the results

In this Appendix we list the proofs of the results shown in this paper.

Proof of Lemma 2. Since $q = 1$ the arrival process $N_0 = \{N_0(t) : t \geq 0\}$ of non repairable items is the zero process and so by relation (13) we obtain introducing

$$k(T) := \int_0^T e^{-\delta u} \lambda(u) c_a(u) du \tag{81}$$

that

$$C(x,\tau) = c_{scr}x + \mathbb{E}\left(\int_0^\tau e^{-\delta u}[c_{se} + c_{re} - c_a(u)]\lambda(u)du\right) + x(h - \delta c_{scr})\mathbb{E}\left(\int_0^\tau e^{-\delta u}du\right) + k(T). \tag{82}$$

This shows for every $\tau \in \mathbb{F}$ that $\inf_{x\in\mathbb{Z}_+}\{c(x) + C(x,\tau)\} = \gamma(\tau) + \mathbb{E}\left(\int_0^\tau e^{-\delta u}[c_{se} + c_{re} - c_a(u)]\lambda(u)du\right) + k(T)$ with the value $\gamma(\tau)$ given by

$$\gamma(\tau) := \inf_{x\in\mathbb{Z}_+}\left\{c(x) + c_{scr}x + x(h - \delta c_{scr})\mathbb{E}\left(\int_0^\tau e^{-\delta u}du\right)\right\}.$$

Since $h - \delta c_{scr} \geq 0$ we obtain by our remark after relation (4) that $\gamma(\tau) = 0$ and so

$$\upsilon(P) = \inf_{\tau\in\mathbb{F}}\left\{\mathbb{E}\left(\int_0^\tau e^{-\delta u}[c_{se} + c_{re} - c_a(u)]\lambda(u)du\right)\right\} + k(T).$$

Since the integrand function in the above optimization problem does not depend on the stopping time τ it is easy to see that an optimal switching time is given by some deterministic stopping time $\tau \in \mathbb{F}_0$ and so

$$\inf_{\tau\in\mathbb{F}}\left\{\mathbb{E}\left(\int_0^\tau e^{-\delta u}[c_{se} + c_{re} - c_a(u)]\lambda(u)du\right)\right\} = \inf_{\tau\in\mathbb{F}_0}\left\{\int_0^\tau e^{-\delta u}[c_{se} + c_{re} - c_a(u)]\lambda(u)du\right\}.$$

We know that the function c_a is non-increasing and by standard first order condition arguments the optimal solution of the above optimization problem is given by ς_1 and no ordering. This has the objective value $\int_0^T e^{-\delta u}\lambda(u)\min\{c_{se} + c_{re}, c_a(u)\}du$ and we have shown the result of Lemma 2.

□

Proof of Lemma 3. Since $c_{csr} \geq 0$ we obtain that all operational costs components mentioned in the beginning of subsection 2.3 are non-negative and so we obtain $C(x,\tau) \geq 0$ for every $x \geq 0$ and $\tau \in \mathbb{F}$. Since $c_{scr} \geq 0$ and hence c is increasing and the vector $(0,0)$ is a feasible solution this implies that for every $x \geq x_U + 1$

$$c(x) + C(x,\tau) \geq c(x) > \int_0^T e^{-\delta u}\lambda(u)c_a(u)du = C(0,0) \geq \upsilon(P) \tag{83}$$

showing the desired result. □

Proof of Lemma 5. Since by Lemma 4 we know that the objective function is bounded below by the function $L(x,\tau)$ and $x \to L(x,\tau)$ is increasing for every $\tau \in \mathbb{F}$, we obtain

$$\upsilon(P) = \inf_{x\in\mathbb{Z}_+,\tau\in\mathbb{F},0\leq\tau\leq T}\{c(x) + C(x,\tau)\} \geq \inf_{\tau\in\mathbb{F},0\leq\tau\leq T}\{L(0,\tau)\}. \tag{84}$$

Similar as in the proof of Lemma 2 it follows using relation (22) that

$$\begin{aligned}\inf_{\tau\in\mathbb{F},0\leq\tau\leq T}\{L(0,\tau)\} &= \inf_{\tau\in\mathbb{F}_0}\left\{\int_0^\tau e^{-\delta u}\lambda(u)[c_{se} + qc_{re} - c_a(u)]du\right\} \\ &= \int_0^{\zeta_q} e^{-\delta u}\lambda(u)[c_{se} + qc_{re} - c_a(u)]du\end{aligned}$$

with ζ_q given in relation (18) and we have verified the result. □

Proof of Lemma 6. It follows by relations (20) for $c_{scr} \geq 0$ and (22) for $c_{csr} \leq 0$ that

$$L(x,\tau) \geq q(x) + \mathbb{E}\left(\int_0^{\tau} e^{-\delta u}\lambda(u)[c_{se}+qc_{re}-c_a(u)]du\right) + \int_0^T e^{-\delta u}\lambda(u)c_a(u)du$$

with $q(x) := c(x) + \min\{c_{scr}, 0\}x$. Since $c_{ap}(u) \geq c_{se}$ for every $0 \leq u \leq T$ we may apply Lemma 4 and so for every $x \in \mathbb{Z}_+$ we obtain by the previous inequality

$$\inf_{\tau\in\mathbb{F},0\leq\tau\leq T}\{c(x)+C(x,\tau)\} \geq q(x) + \inf_{\tau\in\mathbb{F},0\leq\tau\leq T}\left\{\mathbb{E}\left(\int_0^{\tau} e^{-\delta u}\lambda(u)[c_{se}+qc_{re}-c_a(u)]du\right)\right\} + k(T)$$

with $k(T)$ listed in relation (81). This implies using the same arguments as in Lemma 5 that

$$\inf_{\tau\in\mathbb{F},0\leq\tau\leq T}\{c(x)+C(x,\tau)\} \geq q(x) + \int_0^T e^{-\delta u}\lambda(u)\min\{c_{se}+qc_{re}, c_a(u)\}du. \tag{85}$$

Hence it follows for every $x \geq \bar{x}_U$ that

$$\inf_{\tau\in\mathbb{F},0\leq\tau\leq T}\{c(x)+C(x,\tau)\} > \int_0^T e^{-\delta u}\lambda(u)c_a(u)du = C(0,0) \geq \upsilon(P)$$

and the result is verified. □

Proof of Lemma 7. It is sufficient to give the proof of the first result only. The second claim follows replacing f by $-f$. Since f is non-positive it is obvious that the function G is non-increasing. To show that the function G is discrete convex, we note by Doob's stopping theorem for any $0 < \tau \leq T$

$$\mathbb{E}\left(\int_0^{\tau\wedge\sigma_x} f(u)\mu(u)du\right) = \mathbb{E}\left(\int_0^{\tau\wedge\sigma_x} f(u)dN(u)\right) = \mathbb{E}\left(\sum_{k=1}^x f(\sigma_k)1_{\{\sigma_k\leq\tau\}}\right).$$

with σ_k the hitting time at level k of the nonhomogeneous Poisson process N. Since σ_k has a continuous cdf this yields

$$\mathbb{E}\left(\int_0^{\tau\wedge\sigma_x} f(u)\mu(u)du\right) = \mathbb{E}\left(\sum_{k=1}^x f(\sigma_k)1_{\{\sigma_k<\tau\}}\right)$$

and hence for every $x \in \mathbb{Z}_+$ it follows

$$\Delta_x G(x) := G(x+1) - G(x) = \mathbb{E}\left(f(\sigma_{x+1})1_{\{\sigma_{x+1}<\tau\}}\right). \tag{86}$$

Using $\sigma_{x+1} \leq \sigma_{x+2}$ and hence $1_{\{\sigma_{x+1}<\tau\}} \geq 1_{\{\sigma_{x+2}<\tau\}}$ and f non-decreasing and non-positive on $[0,\tau)$ we obtain

$$f(\sigma_{x+1})1_{\{\sigma_{x+1}<\tau\}} \leq f(\sigma_{x+2})1_{\{\sigma_{x+2}<\tau\}}.$$

This shows applying relation (86) that for every $x \in \mathbb{Z}_+$

$$\Delta G(x) = \mathbb{E}(f(\sigma_{x+1})1_{\{\sigma_{x+1}<\tau\}}) \leq \mathbb{E}(f(\sigma_{x+2})1_{\{\sigma_{x+2}<\tau\}}) = \Delta G(x+1),$$

and we have verified the discrete convexity property. □

Proof of Lemma 8. We will only verify the first part of this result. The proof of the second part is similar. Since $c_{ap}(u) \geq c_{se} - c_{scr}$ for every $u < \tau$ and c_{ap} is non-increasing we obtain that the function $u \mapsto e^{-\delta u}(c_{se} - c_{scr} - c_{ap}(u))$ is non-positive and non-decreasing on $[0,\tau)$. Hence, by Lemma 7 the function

$$x \mapsto \mathbb{E}\left(\int_0^{\tau\wedge\sigma_x} e^{-\delta u}\lambda_0(u)(c_{se}-c_{scr}-c_{ap}(u))du\right)$$

is discrete convex. Since the random function $x \mapsto ((x - N_0(u))^+)$ is also discrete convex and $h - \delta c_{scr} \geq 0$ it follows from relation (13) that the function $x \mapsto C(x,\tau)$ is discrete convex again. Finally, the discrete convexity of c completes the proof. □

Proof of Lemma 9. For any $x \in \mathbb{Z}_+$ using c_a decreasing it follows immediately due to $c_a(0) \leq c_{se} + qc_{re} - (1-q)c_{scr}$ and relation (27) that optimization problem (33) has optimal solution $\tau = 0$. This shows using $x \to c(x) + c_{scr}x$ is increasing that it is optimal in optimization problem (32) not to order and immediately start with the alternative policy with objective value $\int_0^T e^{-\delta u}\lambda(u)c_a(u)du$. □

Proof of Lemma 10. For the selected arrival rate and alternative policy cost function in (35) and (34) it is clear for every $x \in \mathbb{Z}_+$ that the function $\tau \to c(x) + C(x \wedge \sigma_x)$ with $C(x \wedge \sigma_x)$ listed in relation (27) is continuous on $(0,T)$ satisfying both $c(x) + C(x, 0 \wedge \sigma_x) = \lim_{\tau \downarrow 0} c(x) + C(x, \tau \wedge \sigma_x)$ and $c(x) + C(x, T \wedge \sigma_x) = \lim_{\tau \uparrow T} c(x) + C(x, \tau \wedge \sigma_x)$. This shows $\psi(x) = \min_{i=1,...,n} \psi_i(x)$. By relation (27) it follows for every $x \in \mathbb{Z}_+$ that the derivative of the function $c(x) + C(x, \tau \wedge \sigma_x)$ with respect to τ for any $a_i < \tau < a_{i+1}, i = 1, ..., n$ is given by

$$\frac{\partial C}{\partial \tau}(x, \tau \wedge \sigma_x) = e^{-\delta\tau}\left(\lambda_i[c_{se} + qc_{re} - (1-q)c_{scr} - c_i]\mathbb{P}(N_0(\tau) < x) + (h - \delta c_{scr})\mathbb{E}((x - N_0(\tau))^+)\right).$$

Clearly for $c_i \leq c_{se} + qc_{re} - (1-q)c_{scr}$ it follows that $\frac{\partial C}{\partial \tau}(x, \tau \wedge \sigma_x) \geq 0$ and so we may immediately conclude using $c_1 \geq c_2 \geq ... \geq c_n$ that the continuous function $\tau \mapsto c(x) + C(x, \tau \wedge \sigma_x)$ is increasing on $[a_{n^*+1}, T]$ for every x. This shows $\psi(x) = \min_{i=1,...,n^*} \psi_i(x)$ and we have shown the result. □

Proof of Lemma 11. For every $x \in \mathbb{N}$ we obtain by writing out the expectation

$$\mathbb{E}((x - N(t))^+) = x\mathbb{P}(N(t) < x) - \sum_{n=1}^{x-1} n\mathbb{P}(N(t) = n)$$

that

$$\frac{\mathbb{E}((x - N(t))^+)}{\mathbb{P}(N(t) < x)} = x - \frac{\sum_{n=1}^{x-1} n\mathbb{P}(N(t) = n)}{\mathbb{P}(N(t) < x)}. \tag{87}$$

Since the random variable $N(t)$ has a Poisson cdf with parameter $\theta(t)$ (cf.Ross (1997)) it follows

$$\frac{\sum_{n=1}^{x-1} n\mathbb{P}(N(t) = n)}{\mathbb{P}(N(t) < x)} = \frac{\sum_{n=1}^{x-1} \frac{\Theta(t)^n}{(n-1)!}}{\sum_{n=0}^{x-1} \frac{\Theta(t)^n}{n!}} = \Theta(t)(1 - B(x-1, \Theta(t)))$$

and substituting this into relation (87) the result follows. □

Proof of Lemma 12. By the interpretation of $B(x,t)$ and expected service time equals 1 it follows by the formula of Little (cf.?) that $t(1 - B(x,t))$ denotes the long run average number of busy serves in the system in a pure Markovian loss system with arrival rate $\lambda = t > 0$ and departure rate $\mu = 1$. Since each server serves exactly one customer and we consider a pure loss system the long run average number of busy servers equals the long run average number of customers in the system and so

$$t(1 - B(x,t)) = \mathbb{E}(\mathbf{Z}^{(t)}(\infty)) \tag{88}$$

with

$$\mathbf{Z}^{(t)}(\infty) := \lim_{s\uparrow\infty} \frac{1}{s}\int_0^s \mathbf{Z}^{(t)}(v)dv$$

denoting the longrun average number of customers in the Markovian loss system in the equilibrium situation having arrival rate t and departure rate 1. Since the Markovian loss system is a birth-death process we obtain by Proposition 4.2.10 of Stoyan and Daley (1983) that

$$\mathbf{Z}^{(t)}(\infty) \geq_d \mathbf{Z}^{(s)}(\infty)$$

for $t > s$ with $\geq_d$ denoting first order dominance and this shows that $\mathbb{E}(\mathbf{Z}^{(t)}(\infty)) \geq \mathbb{E}(\mathbf{Z}^{(s)}(\infty))$. Hence by (88) it follows that the function $t(1 - B(x,t)$ is increasing in t. Since $t(1 - B(x,t))$ represents the number of busy servers and so for the arrival rate going to infinity on average all x servers will be busy it follows that $\lim_{t\uparrow\infty} t(1 - B(t,x)) = x$. □

Proof of Lemma 13. Applying relation (47) it is sufficient to verify for any $i \leq n_0^*$ that

$$\psi_i(x) = \min\{c(x) + C(x, a_i \wedge \sigma_x), c(x) + C(x, a_{i+1} \wedge \sigma_x)\}.$$

By Lemma 12 and using $h - \delta c_{scr} \geq 0$ and Λ_0 increasing it follows that the function κ_i listed in (40) is decreasing. This shows for $\kappa_i(x, a_i) \leq 0$ that by (41) the partial derivative $\frac{\partial C}{\partial \tau}(x, \tau \wedge \sigma_x)$ is non-positive on (a_i, a_{i+1}) and hence the minimum is attained at a_{i+1}. Similar for $\kappa_i(x, a_i) > 0$ we distinguish the mutually exclusive cases $\kappa_i(x, a_{i+1}) > 0$ and $\kappa_i(x, a_{i+1}) \leq 0$. For the first subcase the partial derivative $\frac{\partial C}{\partial \tau}(x, \tau \wedge \sigma_x)$ is non-negative on (a_i, a_{i+1}) and so the minimum is attained at a_i. For the second subcase we conclude using again a standard calculus argument as before that the minimum is either attained at a_i or a_{i+1} showing the desired result. □

Proof of Lemma 14. For $x = 0$ and $i = 1, ..., n+1$ it follows by relation (27) that $C(0, a_i \wedge \sigma_0) = \int_0^T e^{-\delta u} \lambda(u) c_a(u) du$ and we obtain the result in relation (51) using (34) and (35). To verify relation (52) we first observe for every $x \in \mathbb{Z}_+$ that

$$\mathbb{E}((x + 1 - N_0(u))^+) - \mathbb{E}(x - N_0(u))^+) = \mathbb{P}(N_0(u) \leq x)$$

Applying again relation (27) this implies for $i = 2, ..., n+1$ and $x \in \mathbb{Z}_+$

$$\Delta C(x, a_i \wedge \sigma_x) = \begin{cases} c_{scr} + \int_0^{a_i} e^{-\delta u} \lambda(u)[c_{se} + q c_{re} - (1-q) c_{scr} - c_a(u)] \mathbb{P}(N_0(u) = x) du \\ \\ +(h - \delta c_{scr}) \int_0^{a_i} e^{-\delta u} \mathbb{P}(N_0(u) \leq x) du. \end{cases} \tag{89}$$

Substituting now into relation (89) the particular choice of the arrival rate and the alternative policy cost function given in (34) and (35) yields

$$\Delta C(x, a_i \wedge \sigma_x) = \begin{cases} c_{scr} + \sum_{j=1}^{i-1} (c_{se} + q c_{re} - (1-q) c_{scr} - c_j) \lambda_j \int_{a_j}^{a_{j+1}} e^{-\delta u} \mathbb{P}(N_0(u) = x) du \\ \\ +(h - \delta c_{scr}) \sum_{j=1}^{i-1} \int_{a_j}^{a_{j+1}} e^{-\delta u} \mathbb{P}(N_0(u) \leq x) du. \end{cases} \tag{90}$$

To simplify the above expression we observe, since the non-homogenous Poisson process N_0 has the constant arrival rate $(1-q)\lambda_j$ on the interval (a_j, a_{j+1}), that by Lemma 21 using $\rho(u) = e^{-\delta u}$ and applying relation (90) that for $i = 2, ..., n+1$

$$\Delta C(x, a_i \wedge \sigma_x) = \begin{cases} c_{scr} + \sum_{j=1}^{i-1} \left(\frac{c_{se} + q c_{re} - (1-q) c_{scr} - c_j}{1-q}\right) [e^{-\delta a_j} \mathbb{P}(N_0(a_j) \leq x) - e^{-\delta a_{j+1}} \mathbb{P}(N_0(a_{j+1}) \leq x)] \\ \\ + \sum_{j=1}^{i-1} [h + \delta \left(\frac{c_j - c_{se} - q c_{re}}{1-q}\right)] \int_{a_j}^{a_{j+1}} e^{-\delta u} \mathbb{P}(N_0(u) \leq x) du. \end{cases} \tag{91}$$

and this shows the desired result. □

Proof of Lemma 15. To give a proof of Lemma 15 we first show by induction that the sequence $\alpha_{kj}, k \in \mathbb{Z}_+$ given by

$$\alpha_{kj} := \int_{a_j}^{a_{j+1}} e^{-\delta u} \mathbb{P}(N_0(u) = k) du \tag{92}$$

satisfies

$$\alpha_{kj} = \frac{1}{\delta + (1-q)\lambda_j} \sum_{m=0}^{k} \left(\frac{(1-q)\lambda_j}{\delta + (1-q)\lambda_j}\right)^m \theta_j(k-m) \tag{93}$$

with

$$\theta_j(k) := e^{-\delta a_j} \mathbb{P}(N_0(a_j) = k) - e^{-\delta a_{j+1}} \mathbb{P}(N_0(a_{j+1}) = k), k \in \mathbb{Z}_+$$

Introducing the functions $f:(0,T]\to\mathbb{R}$ and $g_k:(0,T]\to\mathbb{R}$, $k\in\mathbb{Z}_+$ given by $f(u):=e^{-(\delta u+(1-q)\Lambda(u))}$ and $g_k(u):=\frac{((1-q)\Lambda(u))^k}{k!}$ it follows for every $k\in\mathbb{Z}_+$ and $a_j<u<a_{j+1}$ that

$$(fg_k)(u)=e^{-\delta u}\mathbb{P}(N_0(u)=k),(fg'_{k+1})(u)=(1-q)\lambda_j e^{-\delta u}\mathbb{P}(N_0(u)=k) \tag{94}$$

Since

$$f'(u)=-(\delta+(1-q)\lambda_j)f(u) \tag{95}$$

for every $a_j<u<a_{j+1}$ we obtain

$$\int_{a_j}^{a_{j+1}} f(u)du=\frac{f(a_j)-f(a_{j+1})}{\delta+(1-q)\lambda_j}=\frac{\theta_j(0)}{\delta+(1-q)\lambda_j}$$

and this verifies relation (93) for $k=0$. To show relation (93) for $k+1$ we assume that relation (93) holds for k. By partial integration and relations (94) and (95) we conclude

$$-(\delta+(1-q)\lambda_j)\alpha_{k+1j}=-(\delta+(1-q)\lambda_j)\int_{a_j}^{a_{j+1}} f(u)g_{k+1}(u)du=-\theta_j(k+1)-(1-q)\lambda_j\alpha_{kj}$$

This shows that

$$\begin{aligned}\alpha_{k+1j} &= \tfrac{(1-q)\lambda_j}{\delta+(1-q)\lambda_j}\alpha_{kj}+\tfrac{\theta_j(k+1)}{\delta+(1-q)\lambda_j}\\ &= \tfrac{1}{\delta+(1-q)\lambda_j}\textstyle\sum_{m=0}^{k}\left(\tfrac{(1-q)\lambda_j}{\delta+(1-q)\lambda_j}\right)^{m+1}\theta_j(k-m)+\tfrac{\theta_j(k+1)}{\delta+(1-q)\lambda_j}\\ &= \tfrac{1}{\delta+(1-q)\lambda_j}\textstyle\sum_{m=0}^{k+1}\left(\tfrac{(1-q)\lambda_j}{\delta+(1-q)\lambda_j}\right)^{m}\theta_j(k+1-m)\end{aligned}$$

and we have verified relation (93). Using this result we obtain

$$\begin{aligned}\textstyle\int_{a_j}^{a_{j+1}} e^{-\delta u}\mathbb{P}(N_0(u)\le x)du &= \textstyle\sum_{k=0}^{x}\alpha_{kj}\\ &= \tfrac{1}{\delta+(1-q)\lambda_j}\textstyle\sum_{k=0}^{x}\sum_{m=0}^{k}\left(\tfrac{(1-q)\lambda_j}{\delta+(1-q)\lambda_j}\right)^{m}\theta_j(k-m)\\ &= \tfrac{1}{\delta+(1-q)\lambda_j}\textstyle\sum_{m=0}^{x}\sum_{k=m}^{x}\left(\tfrac{(1-q)\lambda_j}{\delta+(1-q)\lambda_j}\right)^{m}\theta_j(k-m)\\ &= \tfrac{1}{\delta+(1-q)\lambda_j}\textstyle\sum_{m=0}^{x}\left(\tfrac{(1-q)\lambda_j}{\delta+(1-q)\lambda_j}\right)^{m}\sum_{k=0}^{x-m}\theta_j(k)\end{aligned}$$

and this completes the proof. □

Proof of Lemma 16. Introduce for every $\tau\in\mathbb{F}$ the stopping time $\tau_\Theta=d_\Theta(\tau)$ with

$$d_\Theta(r)=\textstyle\sum_{k=0}^{N-2}t_{k+1}1_{[t_k,t_{k+1})}(r)+T1_{[T_{N-1},T]}(r).$$

By its definition we obtain $\tau_\Theta\ge\tau,\tau_\Theta\in\mathbb{F}_\Theta$ and $\mathbb{E}(\tau_\Theta-\tau)\le\Delta(\Theta)$. Applying relation (13) it follows by some standard upper bound arguments applied to each separate term in this relation that for every $x\in\mathbb{Z}_+$

$$\mid C(x,\tau_\Theta)-C(x,\tau)\mid\le f_0(x)\Delta(\Theta)$$

This shows

$$\upsilon(P_\Theta)-\upsilon(P)\le\sup_{x\le x_U}f_0(x)\Delta(\Theta)=f_0(x_U)\Delta(\Theta)$$

and we have verified the result. □

Proof of Lemma 18. By the definition of W_N and W_{N-1} in relation (65) and (66) it is obvious that $W_{N-1}(x) \leq c_{scr}x = W_N(x)$ and so the result is verified for $k = N-1$. Assume now that $W_{k+1}(x) \leq W_{k+2}(x)$ for every $x \leq x_U$ and $k \leq N-2$. Since the arrival process N_0 is a homogeneous Poisson process it follows that $N_0^{(k)} \stackrel{d}{=} N_0^{(k+1)}$ and this shows using $W_{k+1}(x) \leq W_{k+2}(x)$ that

$$\mathbb{E}(W_{k+1}((x - N_0^{(k)}(\Delta))^+) \leq \mathbb{E}(W_{k+2}((x - N_0^{(k+1)}(\Delta))^+).$$

If it can be shown that $\overline{B}_k(x) \leq \overline{B}_{k+1}(x)$ we obtain by relation (66) that

$$\begin{aligned} W_k(x) &= \min\{c_{scr}x, \overline{B}_k(x) + e^{-\delta\Delta}\mathbb{E}(W_{k+1}((x - N_0^{(k)}(\Delta))^+) \\ &\leq \min\{c_{scr}x, \overline{B}_{k+1}(x) + e^{-\delta\Delta}\mathbb{E}(W_{k+1}((x - N_0^{(k+2)}(\Delta))^+)\} \\ &= W_{k+1}(x) \end{aligned}$$

and this shows the result. Hence it is sufficient to verify that $\overline{B}_k(x) \leq \overline{B}_{k+1}(x)$ and using relation (72) and c_a a decreasing function this is easy to verify for penalty costs being constant on $[0, T]$. □

Proof of Lemma 19. It follows for $x \in S_k$ that $W_k(x) = c_{scr}x$. Applying Lemma 18 we obtain

$$c_{scr}x = W_k(x) \leq W_{k+1}(x) = \min\left\{c_{scr}x, \overline{B}_{k+1}(x) + e^{-\delta\Delta}\mathbb{E}(W_{k+2}((x - N_0^{(k+1)}(\Delta))^+)\right\} \leq c_{scr}x.$$

This shows $W_{k+1}(x) = c_{scr}$ and we have verified the desired result. □

Lemma 21. *Let N be a non-homogeneous Poisson process with a continuous arrival intensity function μ and ρ some differentiable function. Then it follows for every $x \in \mathbb{Z}_+$ and $1 \leq j \leq n$ that*

$$\int_{a_j}^{a_{j+1}} \rho(u)\mu(u)\mathbb{P}(N(u) = x)du = \int_{a_j}^{a_{j+1}} \rho'(u)\mathbb{P}(N(u) \leq x)du + \rho(a_j)\mathbb{P}(N(a_j) \leq x) - \rho(a_{j+1})\mathbb{P}(N(a_{j+1}) \leq x).$$

Proof of Lemma 21. It is well known for every $x \in \mathbb{Z}_+$ (Ross (1997)) that the function $\chi(u) := \mathbb{P}(N(u) \leq x)$ is differentiable and satisfies $\chi'(u) = \mu(u)\mathbb{P}(N(u) = x)$. This shows that

$$\begin{aligned} \rho(a_{j+1})\mathbb{P}(N(a_{j+1}) \leq x) - \rho(a_j)\mathbb{P}(N(a_{j+1}) \leq x) &= \int_{a_j}^{a_{j+1}} (\rho\chi)'(u)du \\ &= \int_{a_j}^{a_{j+1}} \rho'(u)\chi(u)du + \int_{a_j}^{a_{j+1}} \rho(u)\chi'(u)du \\ &= \int_{a_j}^{a_{j+1}} \rho'^{(u)}\chi(u)du + \int_{a_j}^{a_{j+1}} \rho(u)\chi'(u)du \\ &= \int_{a_j}^{a_{j+1}} \rho'(u)\mathbb{P}(N(u) \leq x)du \\ &\quad + \int_{a_j}^{a_{j+1}} \rho(u)\mu(u)\mathbb{P}(N(u) = x)du \end{aligned}$$

and we obtain the desired result. □

Stochastic Models in Probability and Statistics
Vol. 2, No. 2, 2025, pp. 199–209. Research Paper

Lifetime behavior of a new discrete time shock model

Hamed Lorvand[†]* **and Reza Farhadian**[‡]

[†] *Department of Mathematical Sciences, Isfahan University of Technology, Isfahan, 84156-83111, Iran*
[‡] *Department of Statistics, Razi University, Kermanshah, Iran*

Email(s): lorvandhamed@iut.ac.ir, farhadian.reza@yahoo.com

Abstract. In this paper, a new discrete-time shock model is investigated. Under the proposed model, system failure occurs when the inter-arrival time between successive shocks falls below a random critical threshold, denoted by Δ. This model can be viewed as a randomized-threshold extension of the classical delta shock model. Assuming that Δ follows a uniform distribution, we derive the probability mass function of the system's lifetime and analyze the mean time to failure. Furthermore, the lifetime behavior of the system is examined in a Markovian environment. Illustrative numerical results are presented for both the probability mass function of the system's lifetime and the mean time to failure.

Keywords: Discrete time shock model, δ-shock model, Inter-arrival time, Random threshold.

1 Introduction

Shock models are valuable tools for analyzing systems subjected to random external shocks. Understanding how a system behaves under such shocks is essential; consequently, shock models have attracted considerable attention in applied probability, reliability theory, and engineering. Common types of shock models include cumulative shock models, extreme shock models, and run shock models. In a cumulative shock model, system failure results from the accumulated effect of shocks; in an extreme shock model, failure is caused by a single large shock; and in a run shock model, failure is determined by the occurrence of a specified number of consecutive shocks within a given range. For more on these, see, e.g., Gut (1990), Anderson (1998), Mallor and Omey (2001), and Sumita and Shanthikumar (1985).

There are other shock models that have been developed in recent years. The so-called δ-*shock model* is one of them that has attracted more attention. According to the classical δ-shock model, if the inter-arrival time between two consecutive shocks falls below a critical threshold δ, the system fails. Thus, the system's lifetime is defined as $T = \sum_{i=1}^{N} X_i$, where $X_1, X_2, \ldots$ are inter-arrival times, and N is the number of shocks until the failure of the system. Thus, $\{N = n\}$ iff $\{X_1 > \delta, \ldots, X_{n-1} > \delta, X_n \leq \delta\}$. The δ-shock model was first introduced by Li et al. (1999), after which it was widely studied by many researchers; see, e.g., Li et al. (1999), Xu and Li (2004), Li and Kong (2007), Li and Zhao (2007), Wang and Zhang (2007), and Ma and Li (2010). Many extensions and generalizations have been proposed for the δ-shock model. We briefly review some of them below. Eryilmaz (2012) proposed a generalized δ-shock model, in which the system fails when k consecutive inter-arrival times are less than a threshold δ. Eryilmaz (2013) studied a discrete time version of the δ-shock model, where the shocks occur according to a

*Corresponding author
Received: 6 May 2025 / Accepted: 9 December 2025

 https://smps.um.ac.ir

binomial process, i.e. the inter-arrival times between successive shocks follow a geometric distribution. Eryilmaz and Bayramoglu (2014) studied the lifetime behavior of the δ-shock model and the censored δ-shock model under the assumption that shocks arrive according to a renewal process with uniformly distributed inter-arrival times. Tuncel and Eryilmaz (2018) investigated the survival function and the mean lifetime of the system failure under the δ-shock model, considering the proportional hazard rate model. Poursaeed (2019) studied the reliability analysis of an extension of the discrete time version of the δ-shock model by considering two different critical thresholds and a probable failure region. Lorvand et al. (2020) discussed a mixed δ-shock model for multi-state systems by assuming a renewal process of shocks where the system fails in three specific states. Goyal et al. (2022) introduced a general δ-shock model in which the recovery time depends on both the inter-arrival time and the magnitude of the shocks. Finkelstein and Cha (2024) revisited the classical δ-shock model and generalized it to the case of renewal processes of external shocks with arbitrary inter-arrival times and arbitrary distribution of the recovery parameter δ.

In this paper, we study a new discrete time shock model in which the system fails if the inter-arrival time between two successive shocks is less than or equal to a random critical threshold, denoted by Δ. This model can be regarded as a randomized-threshold extension of the classical δ-shock model. The motivation for introducing a random threshold stems from the fact that, in many real-world systems, assuming a fixed threshold is unrealistic. Such situations may arise due to heterogeneity and uncertainty in the strength, tolerance, or operating conditions of systems subjected to shock environments.
Throughout this paper, we study the lifetime of the system under the proposed model. In particular, we derive the probability mass function of the system's lifetime and the mean time to failure of the system. Furthermore, the lifetime behavior of the system in a Markovian environment is investigated.

The remainder of this paper is organized as follows. Section 2 introduces the model and the underlying assumptions. The probabilistic behavior of the system's lifetime under the proposed model is derived in Section 3. A Markovian process approach is presented in Section 4. Finally, Section 5 concludes the paper.

2 Description of the model

Assume that a system is subject to a sequence of external random shocks. Let X_i be the inter-arrival time between the ith and $(i+1)$th shocks for $i=1,2,\ldots$, and let also inter-arrival times $X_1,X_2,\ldots$ take nonnegative integer values and are independent and identically distributed by an arbitrary discrete probability distribution with the probability mass function (pmf) $\Pr(X=x)$, where X is a generic random variable of X_i's. The performance of the system is such that for a random critical threshold Δ, which is independent of X_i's, if $X_n \leq \Delta$, the system fails. Thus, the lifetime of the system is defined as follows:

$$T_\Delta = \sum_{i=1}^{N} X_i,$$

where

$$\{N=n\} \Leftrightarrow \{X_1>\Delta, X_2>\Delta, \ldots, X_{n-1}>\Delta, X_n \leq \Delta\}. \tag{1}$$

By considering an appropriate probability distribution for Δ, the relationship of the new model to the classical δ-shock model can be expressed as in the following remark:

Remark 1. *If Δ has a degenerate distribution at point δ, i.e., $\Pr(\Delta=\delta)=1$, the model reduces to the classical δ-shock model*

The proposed model can be applied to a variety of real-world systems, an example of which is described below. Many electrical components used in electronic devices may overheat and fail when exposed to electric shocks. Examples of such components include capacitors, transistors, and CPUs, which can overheat due to electrical disturbances such as high voltage and low current.

Consider a specific component operating in an environment with instantaneous temperature changes that heats up with each shock arriving over discrete time periods. If the next shock reaches the component shortly after the previous one, it may not have sufficient time to cool down and may therefore fail due to overheating. If the heat generated by each shock is the same and the ambient temperature remains constant, a fixed critical threshold for the inter-arrival time between two consecutive shocks could be determined. However, because the ambient temperature varies over time, the cooling time of the component is also affected. As a result, assuming a fixed critical threshold becomes unrealistic, and the critical threshold is more appropriately modeled as a random variable. Therefore, in such scenarios, using the proposed model can provide a more realistic assessment of system survival.

3 Probability behavior of the system's lifetime

3.1 Properties of the stopping time N

In the following theorem, we derive the pmf of random variable N under the model described in Section 2. We obtain the pmf of N in the following theorem, assuming that the random critical threshold Δ is uniformly distributed.

Theorem 1. *Consider the model described in Section 2. If Δ has support in $\{0,1,2,\ldots,a\}$ with the pmf $\Pr(\Delta=\delta)=\frac{1}{a+1}$ for $\delta\in\{0,1,2,\ldots,a\}$, then the pmf of N is*

$$\Pr(N=n)=\frac{1}{a+1}\sum_{\delta=0}^{a}(1-\Pr(X\le\delta))^{n-1}\Pr(X\le\delta).$$

Proof. Since X_i's are independent of Δ, by using the definition of N (see (1)) we have

$$\Pr(N=n)=\Pr(X_1>\Delta,X_2>\Delta,\ldots,X_{n-1}>\Delta,X_n\le\Delta).$$

Besides, since $\Pr(\Delta=\delta)=\frac{1}{a+1}$ for $\delta=0,1,2,\ldots,a$, then

$$\begin{aligned}\Pr(N=n)&=\sum_{\delta=0}^{a}\Pr(X_1>\Delta,X_2>\Delta,\ldots,X_{n-1}>\Delta,X_n\le\Delta,\Delta=\delta)\\&=\sum_{\delta=0}^{a}\Pr(X_1>\delta,X_2>\delta,\ldots,X_{n-1}>\delta,X_n\le\delta,\Delta=\delta)\\&=\sum_{\delta=0}^{a}\Pr(X_1>\delta,X_2>\delta,\ldots,X_{n-1}>\delta,X_n\le\delta)\Pr(\Delta=\delta)\\&=\sum_{\delta=0}^{a}(\Pr(X>\delta))^{n-1}\Pr(X\le\delta)\left(\frac{1}{a+1}\right)\\&=\frac{1}{a+1}\sum_{\delta=0}^{a}(1-\Pr(X\le\delta))^{n-1}\Pr(X\le\delta).\end{aligned}$$

This completes the proof. □

Corollary 1. *Let the inter-arrival time X have a geometric distribution with mean $\frac{1}{p}$ for $0<p<1$. Then, the pmf of N is*

$$\Pr(N=n)=\frac{1}{a+1}\sum_{\delta=0}^{a}\left(q^{\delta+1}\right)^{n-1}(1-q^{\delta+1}),$$

where $q=1-p$.

Proof. We have $\Pr(X=x)=p(1-p)^{x-1}$ for $x=1,2,\ldots$ and $0<p<1$. Using this in Theorem 1, we have

$$\Pr(N=n)=\frac{1}{a+1}\sum_{\delta=0}^{a}\left(q^{\delta+1}\right)^{n-1}(1-q^{\delta+1}),\qquad q=1-p.$$

Now, we shall show that $\sum_{n=1}^{\infty}\Pr(N=n)=1$. We have

$$\begin{aligned}\sum_{n=1}^{\infty}\Pr(N=n)&=\sum_{n=1}^{\infty}\left(\frac{1}{a+1}\sum_{\delta=0}^{a}\left(q^{\delta+1}\right)^{n-1}(1-q^{\delta+1})\right)\\&=\frac{1}{a+1}\sum_{\delta=0}^{a}\left(\left(1-q^{\delta+1}\right)\sum_{n=1}^{\infty}\left(q^{\delta+1}\right)^{n-1}\right)\\&=\frac{1}{a+1}\sum_{\delta=0}^{a}(1-q^{\delta+1})\frac{1}{1-q^{\delta+1}}\\&=\frac{1}{a+1}\sum_{\delta=0}^{a}1\\&=1.\end{aligned}$$

This completes the proof. □

In the next theorem, we obtain the mean of the random variable N.

Theorem 2. *Consider the model described in Section 2. If Δ has support in $\{0,1,2,\ldots,a\}$ with the pmf $\Pr(\Delta=\delta)=\frac{1}{a+1}$ for $\delta\in\{0,1,2,\ldots,a\}$, then the mean of N is*

$$E(N)=\frac{1}{a+1}\sum_{\delta=0}^{a}\frac{1}{\Pr(X\leq\delta)}.$$

Proof. We have

$$\begin{aligned}E(N)&=\sum_{n=1}^{\infty}\frac{1}{a+1}\sum_{\delta=0}^{a}n\Pr(X\leq\delta)\left(1-\Pr(X\leq\delta)\right)^{n-1}\\&=\frac{1}{a+1}\sum_{\delta=0}^{a}\sum_{n=1}^{\infty}n\Pr(X\leq\delta)\left(1-\Pr(X\leq\delta)\right)^{n-1}\\&=\frac{1}{a+1}\sum_{\delta=0}^{a}\Pr(X\leq\delta)\sum_{n=1}^{\infty}n\left(1-\Pr(X\leq\delta)\right)^{n-1}\\&=\frac{1}{a+1}\sum_{\delta=0}^{a}\Pr(X\leq\delta)\frac{1}{\left(\Pr(X\leq\delta)\right)^{2}}\\&=\frac{1}{a+1}\sum_{\delta=0}^{a}\frac{1}{\Pr(X\leq\delta)}.\end{aligned}$$

The proof is complete. □

3.2 The pmf of the system's lifetime

Under the model described in Section 2, the pmf of the system's lifetime is given below.

Theorem 3. *Consider the model described in Section 2, and let the random critical threshold* Δ *take values from* $\{0,1,2,\dots,a\}$ *with the pmf* $\Pr(\Delta=\delta)=\frac{1}{a+1}$. *If the inter-arrival times follow the geometric distribution with mean* $\frac{1}{p}$, *the pmf of* T_Δ *is*

$$\begin{aligned}\Pr(T_\Delta=t) &= \sum_{\delta=0}^{a}\Pr(T_\Delta=t|\Delta=\delta)\Pr(\Delta=\delta)\\ &= \sum_{\delta=0}^{a}\sum_{n=2}^{\lfloor\frac{t+\delta+1}{\delta+2}\rfloor}\Pr(\Delta=\delta)\left[\binom{t-(n-2)\delta-(n-1)}{n-1}-\binom{t-(n+1)\delta-(n-2)}{n-1}\right]p^n(1-p)^{t-n}.\end{aligned}$$

Proof. The following pattern shows the lifetime of the system. In that way, 1 means that a shock has occurred, and 0 means that no shock has occurred.

$$\underbrace{\underbrace{0\dots0}_{y_1\geq 0}1\underbrace{0\dots0}_{y_2>\Delta}1\underbrace{0\dots0}_{y_3>\Delta}1\dots1\underbrace{0\dots0}_{y_{n-1}>\Delta}1\underbrace{0\dots0}_{y_n\leq\Delta}1}_{t}.$$

Indeed, the number of 0s until a 1 occurs is considered inter-arrival time. The total possible states of the above pattern are equal to the number of integer solutions of the following equation:

$$\underset{y_1\geq 0,y_2>\Delta,\dots,y_{n-1}>\Delta,y_n\leq\Delta}{y_1+y_2+\dots+y_{n-1}+y_n=t-n},$$

where n is the number of 1's. Then we have

$$\underset{y_1\geq 0,y_2>\Delta+1,\dots,y_{n-1}>\Delta+1,y_n\leq\Delta}{y_1+y_2+\dots+y_{n-1}=t-n.}$$

This is equivalent to the equation

$$\underset{y_1\geq 0,y_2\geq\Delta+1,\dots,y_{n-1}\geq\Delta+1,y_n\leq\Delta}{y_1+y_2-(\Delta+1)+\dots+y_{n-1}-(\Delta+1)+y_n=t-n-(n-2)(\Delta+1),}$$

that is,

$$\underset{y_1\geq 0,y_2'\geq 0,\dots,y_{n-1}'\geq 0,y_n\leq\Delta}{y_1+y_2'+\dots+y_{n-1}'+y_n=t-n-(n-2)\Delta-(n-2).}$$

The number of integer solutions of the above equation is equal to the difference between the number of integer solutions of the equations

$$\underset{y_1\geq 0,y_2'\geq 0,\dots,y_{n-1}'\geq 0,y_n\geq 0}{y_1+y_2'+\dots+y_{n-1}'+y_n=t-n-(n-2)\Delta-(n-2),} \tag{2}$$

and

$$\underset{y_1\geq 0,y_2'\geq 0,\dots,y_{n-1}'\geq 0,y_n\geq\Delta+1}{y_1+y_2'+\dots+y_{n-1}'+y_n=t-n-(n-2)\Delta-(n-2).} \tag{3}$$

The number of integer solutions of (2) is

$$\binom{t-(n-2)\Delta-(n-1)}{n-1},$$

and the number of integer solutions of (3) is

$$\binom{t-(n+1)\Delta-(n-2)}{n-1}.$$

This completes the proof. □

In the next theorem, we obtain an explicit formula for the mean lifetime of the system, which defines the system's mean time to failure (MTTF).

Theorem 4. *Consider the model described in Section 2, and let the random critical threshold* Δ *take values from* $\{0,1,2,\ldots,a\}$ *with the pmf* $\Pr(\Delta=\delta)=\frac{1}{a+1}$. *The MTTF of the system is*

$$E(T_\Delta)=\frac{E(X)}{a+1}\sum_{\delta=0}^{a}\frac{1}{\Pr(X\leq\delta)}.$$

Proof. The random variable N is a stopping time for $X_1,X_2,\ldots$, i.e., the event $\{N=n\}$ is independent of inter-arrival times $X_{n+1},X_{n+2},\ldots$ for all integers $n\geq 1$. Hence, by using the well-known Wald's identity, the MTTF of the system can be computed as follows:

$$E(T_\Delta)=E\left(\sum_{i=1}^{N}X_i\right)=E(N)E(X_1).$$

By using Theorem 2 for $E(N)$, the desired result is immediately obtained. □

3.3 Numerical results

In this section, we give some numerical results on the pmf and MTTF of the system's lifetime. To this end, suppose a system is subject to random shocks under the model described in Section 2, and let the inter-arrival times between successive shocks have the geometric distribution with mean $\frac{1}{p}$ for $p>0$. Tables 1 and 2 show the pmf and MTTF of the system's lifetime T_Δ for the cases that the random critical threshold Δ follows the uniform distribution on the set $\{0,1,2,3\}$ and $\{0,1,2,3,4,5\}$, respectively. In both tables, incremental values of 0.2, 0.5, and 0.6 are considered for the parameter p, which is equivalent to a decreasing trend for the mean of inter-arrival times between successive shocks.

Table 1: The pmf and MTTF of T_Δ for when $\Delta\sim Unif\{0,1,2,3\}$.

p	$E(T_\Delta)$	t	$\Pr(T_\Delta=t)$
0.2	2.7801	3	0.0798
		4	0.0591
		5	0.0434
		7	0.0269
0.5	0.3678	3	0.0201
		4	0.0101
		5	0.0065
		7	0.0021
0.6	0.3589	3	0.0021
		4	0.0010
		5	0.0002
		7	0.0001

Table 2: The pmf and MTTF of T_Δ for when $\Delta \sim Unif\{0,1,2,3,4,5\}$.

p	$E(T_\Delta)$	t	$\Pr(T_\Delta = t)$
0.2	2.1571	3	0.0591
		4	0.0487
		5	0.0391
		7	0.0222
0.5	0.7650	3	0.0551
		4	0.0301
		5	0.0011
		7	0.0005
0.6	0.3671	3	0.0014
		4	0.0009
		5	0.0001
		7	0.0000

From Tables 1 and 2, it can be seen that the values of MTTF and pmf decrease with increasing parameter p. This seems reasonable because increasing p leads to a decrease in the mean of inter-arrival times between shocks and, consequently, the probability of occurrence of inter-arrival times in the sets $\{0,1,2,3\}$ and $\{0,1,2,3,4,5\}$ increases, which is equivalent to a decrease in the probability of the system's lifetime T_Δ.

3.4 Estimation problem

This section deals with the estimation of the parameter of the critical threshold distribution. In the model under discussion, we have $\Delta \sim Unif\{0,1,2,\ldots,a\}$ and the inter-arrival times between successive shocks are distributed geometrically with mean $\frac{1}{p}$ for $p > 0$. Assuming that p (and then $q = 1-p$) is known, we seek to obtain an estimator $\hat{a}$ for the parameter a. Note that the parameter p is related to the probability behavior of the random shocks, but the threshold Δ depends on the system's resistance conditions. Thus, when the parameter p is assumed to be known, it means that we have already discovered the probability behavior of random shocks and intend to use inference to understand the resilient behavior of the system through estimating the parameter a of the probability distribution of Δ.

Let $t_1, t_2, \ldots, t_n$ be the observed values of a random sample of the lifetime data based on n independent systems, and let also $\bar{t} = \frac{1}{n}\sum_{i=1}^{n} t_i$ be the sample mean. Using the method of moments and using Theorem 4, we can obtain the moment estimate of the parameter a by solving the following equation:

$$E(T_\Delta) = \frac{\frac{1}{p}}{a+1} \sum_{\delta=0}^{a} \frac{1}{\left(1-q^{\delta+1}\right)} = \bar{t}. \tag{4}$$

Since a and q are positive, we have

$$\sum_{\delta=0}^{a} \frac{1}{\left(1-q^{\delta+1}\right)} = \frac{-\psi_q(a+2)+\psi_q(1)+(a+1)\log(q)}{\log(q)},$$

where $\psi_q(z)$ is the q-analogue of the digamma function, which is called the q-digamma function and is defined as $\psi_q(z) = \frac{1}{\Gamma_q(z)}\frac{\partial \Gamma_q(z)}{\partial z}$, where $\Gamma_q(z)$ is the q-gamma function. For more details about these functions, see, e.g., Singh et al. (2016).

Therefore, Eq. (4) can be written as follows:

$$E(T_\Delta) = \frac{-\psi_q(a+2)+\psi_q(1)}{(a+1)p\log(q)} + \frac{1}{p} = \bar{t}. \tag{5}$$

By solving Eq. (5), the estimate $\hat{a}$ of parameter a is obtained. However, it is clear that solving Eq. (5) is not possible through conventional methods and requires complex numerical calculations.

4 Markovian shock occurrences

In the previous section, the results were obtained based on the assumption that the occurrences of shocks are independent. In this section, we study the behavior of the system's lifetime by considering the shock occurrences as a Markov chain. We consider the two following patterns:

$$1\underbrace{\underbrace{0\ldots0}_{y_1>\Delta}1\underbrace{0\ldots0}_{y_2>\Delta}1\underbrace{0\ldots0}_{y_3>\Delta}1\ldots1\underbrace{0\ldots0}_{y_{n-1}>\Delta}1\underbrace{0\ldots0}_{y_n\leq\Delta}1}_{t}, \tag{6}$$

and

$$\underbrace{\underbrace{0\ldots0}_{y_1\geq 0}1\underbrace{0\ldots0}_{y_2>\Delta}1\underbrace{0\ldots0}_{y_3>\Delta}1\ldots1\underbrace{0\ldots0}_{y_{n-1}>\Delta}1\underbrace{0\ldots0}_{y_n\leq\Delta}1}_{t}. \tag{7}$$

In the next two theorems, we obtain the pmf of the system's lifetime under the above two situations.

Theorem 5. *Consider the model described in Section 2, and let the random critical threshold* Δ *take values from* $\{0,1,2,\ldots,a\}$ *with the pmf* $\Pr(\Delta=\delta)=\frac{1}{a+1}$. *Under the pattern in Eq. (6), the pmf of* T_Δ *is*

$$\Pr(T_\Delta=t)=\sum_{\delta=0}^{a}\Pr(\Delta=\delta)\sum_{n=2}^{\lfloor\frac{t+\delta+1}{\delta+2}\rfloor}\left[\binom{t-(n-1)(\delta+1)-1}{n-1}-\binom{t-n\delta-(n-2)}{n-1}\right]p_1p_{10}^{n-1}p_{00}^{t-2n+2}p_{01}^{n-1},$$

where $p_{ij}=\Pr(I_n=j|I_{n-1}=i)$ *for* $i,j=0,1$, *and*

$$I_n=\begin{cases}1 & \text{if a shock occurs at time } n,\\ 0 & \text{otherwise,}\end{cases}$$

and $p_1=\Pr(I_1=1)$.

Proof. In the pattern (6), we let the number of shocks be n and we have

$$\underset{y_1>\Delta,y_2>\Delta,\ldots,y_{n-1}>\Delta,y_n\leq\Delta}{y_1+y_2+\cdots+y_{n-1}+y_n=t-n.}$$

This equation is equivalent to the equation

$$\underset{y_1\geq\Delta+1,y_2\geq\Delta+1,\ldots,y_{n-1}\geq\Delta+1,y_n\leq\Delta}{y_1+y_2+\cdots+y_{n-1}=t-n}\ . \tag{8}$$

We can rewrite Eq. (8) as follows:

$$\underset{y_1\geq 0,y_2\geq 0,\ldots,y_{n-1}\geq 0,y_n\leq\Delta}{y_1-(\Delta+1)+y_2-(\Delta+1)+\cdots+y_{n-1}-(\Delta+1)+y_n=t-n-(n-1)(\Delta+1),}$$

that is equivalent to the equation

$$\underset{y_1'\geq 0,y_2'\geq 0,\ldots,y_{n-1}'\geq 0,y_n\leq\Delta}{y_1'+y_2'+\cdots+y_{n-1}'+y_n=t-n-(n-1)(\Delta+1).} \tag{9}$$

The number of integer solutions of Eq. (9) is equal to the difference between the number of integer solutions of the following two equations:

$$\underset{y_1'\geq 0, y_2'\geq 0,\ldots,y_{n-1}'\geq 0, y_n\geq 0}{y_1'+y_2'+\cdots+y_{n-1}'+y_n = t-n-(n-1)(\Delta+l)}, \tag{10}$$

and

$$\underset{y_1'\geq 0, y_2'\geq 0,\ldots,y_{n-1}'\geq 0, y_n\geq \Delta}{y_1'+y_2'+\cdots+y_{n-1}'+y_n = t-n-(n-1)(\Delta+1)}. \tag{11}$$

The number of integer solutions of Eq. (10) is

$$\binom{t-(n-1)(\Delta+1)-1}{n-1},$$

and the number of integer solutions of Eq. (11) is

$$\binom{t-n\Delta-(n-2)}{n-1}.$$

Thus, the number of integer solutions of Eq. (9) is

$$\binom{t-(n-1)(\Delta+1)-1}{n-1} - \binom{t-n\Delta-(n-2)}{n-1}.$$

This completes the proof. □

Theorem 6. *Consider the model described in Section 2, and let the random critical threshold* Δ *take values from* $\{0,1,2,\ldots,a\}$ *with the pmf* $\Pr(\Delta=\delta)=\frac{1}{a+1}$*. Under the pattern in Eq. (7), the pmf of* T_Δ *is*

$$\Pr(T_\Delta = t) = \sum_{\delta=0}^{a} \Pr(\Delta=\delta) \sum_{n=2}^{\lfloor \frac{t}{\delta+2}+2 \rfloor} \left[\binom{t-(n-2)(\delta+1)+1}{n-1} - \binom{t-(n-1)(\delta+1)+1}{n-1}\right] p_0 p_{10}^{n-1} p_{00}^{t-2n+1} p_{01}^{n},$$

where $p_{ij} = \Pr(I_n = j | I_{n-1} = i)$ *for* $i,j=0,1$*, and*

$$I_n = \begin{cases} 1 & \textit{if a shock occurs at time } n, \\ 0 & \textit{otherwise,} \end{cases}$$

and $p_0 = \Pr(I_1 = 0)$.

Proof. In the pattern (7), we let the number of shocks is n and we have

$$\underset{y_1\geq 0, y_2>\Delta,\ldots,y_{n-1}>\Delta, y_n\leq \Delta}{y_1+y_2+\cdots+y_{n-1}+y_n = t-n}.$$

This equation is equivalent to the equation

$$\underset{y_1\geq 0, y_2\geq \Delta+1,\ldots,y_{n-1}\geq \Delta+1, y_n\leq \Delta}{y_1+y_2-(\Delta+1)+\cdots+y_{n-1}-(\Delta+1)+y_n = t-n-(n-2)(\Delta+1)},$$

that is,

$$\underset{y_1\geq 0, y_2'\geq 0,\ldots,y_{n-1}'\geq 0, y_n\leq \Delta}{y_1+y_2'+\cdots+y_{n-1}'+y_n = t-n-(n-2)(\Delta+1)}.$$

The number of integer solutions of the above equation is

$$\binom{t-(n-2)(\Delta+1)+1}{n-1} - \binom{t-(n-1)(\Delta+1)+1}{n-1}.$$

The proof is complete. □

5 Conclusions and future directions

We have introduced a new discrete-time shock model in which the system fails whenever the inter-arrival time between successive shocks is less than or equal to a random critical threshold Δ. In fact, this model is a randomized-threshold version of the classical $\boldsymbol{\delta}$-shock model. Moreover, we have shown that when Δ is supported on the set $\{\delta\}$ with $\delta > 0$, that is, when Δ has a degenerate distribution, the proposed model reduces to the classical $\boldsymbol{\delta}$-shock model. We have derived the pmf of both the system's stopping time and the system's lifetime under the proposed model by assuming that the inter-arrival times follow a geometric distribution and that Δ is uniformly distributed. We have also considered the case in which shock occurrences are dependent within a Markovian framework.

From a theoretical perspective, introducing a random threshold into $\boldsymbol{\delta}$-shock modeling provides a more advanced framework for studying this class of problems and enables the modeling of systems with complex failure mechanisms. Therefore, based on this study, future research can be directed toward more complex shock environments in which system operating conditions are more intricate than before. One important source of this complexity may be the randomness of critical thresholds against shocks. This idea can also be extended to settings in which both time and threshold are modeled as continuous random variables.

Acknowledgements

The authors would like to thank the Editor and two anonymous reviewers for their valuable comments and suggestions that helped improve the original version of the paper.

References

Anderson, K. K., (1988). A note on cumulative shock models. *Journal of Applied Probability*, **25**, 220–223.

Eryilmaz, S. (2012). Generalized $\boldsymbol{\delta}$-shock model via runs. *Statistics and Probability Letters*, **82**, 326–331.

Eryilmaz, S. (2013). On the lifetime behavior of a discrete time shock model. *Journal of Computational and Applied Mathematics*, **237**, 384–388.

Eryilmaz, S and Bayramoglu, K. (2014). Life behavior of $\boldsymbol{\delta}$-shock models for uniformly distributed inter-arrival times. *Statistical Papers*, **55**, 841–852.

Finkelstein, M. and Cha, J. H. (2024). Discussing some approaches to delta-shock modeling. *TOP Transactions in Operations Research*, **32**, 245–262.

Goyal, D., Hazra, N. K., and Finkelstein, M. (2022). On the general $\boldsymbol{\delta}$-shock model. *TEST*, **31**, 994–1029.

Gut, A. (1990). Cumulative shock models. *Advances in Applied Probability*, **22**, 504–507.

Li, Z.H., Chan, L.Y. and Yuan, Z. X. (1999). Failure time distribution under a $\boldsymbol{\delta}$-shock model and its application to economic design of system. *International Journal of Reliability, Quality and Safety Engineering*, **6**, 237–247.

Li, Z. H., Huang, B. S. and Wang, G. J. (1999). Life distribution and its properties of shock models under random shocks. *Journal of Lanzhou University*, **35**, 1–7.

Li, Z.H. and Kong, X.B. (2007). Life behavior of $\boldsymbol{\delta}$-shock model. *Statistics and Probability Letters*, **77**, 577–587.

Li, Z.H. and Zhao, P. (2007). Reliability analysis on the δ-shock model of complex systems. *IEEE Transactions on Reliability*, **56**, 340–348.

Lorvand, H., Nematollahi, A. and Poursaeed, M. H. (2020). Life distribution properties of a new δ-shock model. *Communications in Statistics - Theory and Methods*, **49**(12), 3010–3025.

Ma, M. and Li, Z. (2010). Life behavior of censored δ-shock model. *Indian Journal of Pure and Applied Mathematics*, **41**,401–420.

Mallor, F. and Omey E. (2001). Shocks, runs and random sums. *Journal of Applied Probability*, **38**, 438–448.

Poursaeed, M. H. (2019). A generalization of δ-shock model. *Journal of Advanced in Mathematical Modeling*, **9**, 58–70.

Singh, V. K., Srivastava, H.M., Venturino, E., Resch, M. and Gupta, V. (2015). *Modern Mathematical Methods and High Performance Computing in Science and Technology*, Springer, Singapore.

Sumita, U. and Shanthikumar, J.G. (1985). A class of correlated cumulative shock models. *Advances in Applied Probability*, **17**, 347–366.

Tuncel, A. and Eryilmaz, S. (2018). System reliability under δ-shock model. *Communications in Statistics - Theory and Methods*, **47**, 4872–4880.

Wang, G. J. and Zhang, Y.L. (2007). A shock model with two-type failures and optimal replacement policy. *International Journal of Systems Science*, **36**, 209–214.

Xu, Z.Y. and Li, Z. H. (2004). Statistical inference on d shock model with censored date. *Chinese Journal of Applied Probability and Statistics*, **20**, 147–153.

Stochastic Models in Probability and Statistics
Vol. 2, No. 2, 2025, pp. 211–219. Research Paper

Neutrosophic ZLindley distribution with applications

Lazhar Benkhelifa*

Department of Mathematics, Mohamed Khider University, Biskra, Algeria

Email(s): lazhar.benkhelifa@univ-biskra.dz

Abstract. This paper introduces the neutrosophic ZLindley distribution, extending the classical Lindley distribution within a neutrosophic framework to more effectively handle datasets characterized by uncertainty, imprecision, and incompleteness. We derive several properties of the proposed model and estimate its neutrosophic parameter using the method of maximum likelihood. A simulation study is performed to evaluate the estimator's performance. Finally, the advantages of the proposed model are demonstrated through applications to real datasets.

Keywords: Data analysis, Imprecise data, Maximum likelihood method, Neutrosophic random variable, ZLindley distribution.

1 Introduction

When data or portions of it are uncertain to varying degrees, the neutrosophic theory presents a powerful framework to handle uncertainty. Although certainty is highlighted as a basic requirement in classical mathematics, real-world data usually has high levels of fuzziness, uncertainty, and incompleteness, often surpassing the existence of fully determinate information. To fill this gap, Smarandache (2014) suggested the notion of neutrosophic statistics as a logical extension of classical statistics. The purpose of this expansion is to manage imprecise or uncertain data and their accompanying statistical probability distributions by describing them via sets of values that approximate the original idealized precise values. In these situations, the usefulness of neutrosophic statistical techniques is especially clear, since they allow us to analyze and arrange neutrosophic data in ways that more accurately and nuancedly reveal underlying patterns than are possible with traditional statistical techniques.

In recent years, researchers have developed neutrosophic analogues of numerous classical probability distributions. For example, Alhabib et al. (2018) introduced neutrosophic versions of the Poisson, exponential, and uniform distributions. Alhasan and Smarandache (2019) formulated the neutrosophic Weibull distribution. Khan Sherwani el al. (2021) introduced the neutrosophic beta distribution. Alduais and Khan (2025) developed the neutrosophic extensions of the log-normal distribution. Bashir el al. (2024) presented a modified neutrosophic log-logistic distribution. Bashir el al. (2025) proposed the neutrosophic Lindley distribution. Alduais and Khan (2025) explored the neutrosophic gamma distribution. Razmkhah el al. (2026) investigated the neutrosophic Birnbaum-Saunders distribution.

*Corresponding author
Received: 25 September 2025 / Accepted: 11 December 2025
DOI: 10.22067/smps.2025.95633.1053

Some existing methods for interval data involve calculating classical statistical measures on the lower and upper bounds of the data intervals. This separate calculation and combination of bounds can neglect the internal details and variability within the interval, resulting in a loss of important information. For this reason, Razmkhah el al. (2026) propose a new statistical method for handling interval data that directly uses the entire data interval. This approach allows for the preservation of the complete information contained within the data interval, leading to richer datasets for analysis and offering more accurate and reliable statistical results.

In this paper, we propose the neutrosophic ZLindley distribution (NZLD). The rest of the paper is organized as follows: Section 2 introduces the NZLD. Some statistical properties are presented in Section 3. Section 4 describes the estimation procedure for the neutrosophic parameter. A simulation study is carried out in Section 5. In Section 6, real datasets are analyzed to demonstrate the applicability of the proposed distribution. Finally, a brief conclusions are provided in Section 7.

2 Neutrosophic ZLindley distribution

The increasing complexity of real-life data demands more flexible and robust probability models. Traditional parametric models, like exponential, Weibull, and normal distributions, often fail to capture the nuanced behavior of modern survival or reliability data. To address these challenges, researchers and practitioners have recently proposed several models. Saaidia et al. (2024) introduced the ZLindley and studied some of its properties. A random variable X follows the ZLindley distribution (ZLD) if its cumulative distribution function (cdf) is

$$F(x,\theta) = 1 - \left(1 + \frac{\theta x}{2(\theta+1)}\right) e^{-\theta x}, \quad x > 0,$$

where $\theta > 0$ is the scale parameter. The corresponding probability density function (pdf) of X is given by

$$f(x,\theta) = \frac{\theta(1+2\theta+\theta x)}{2(\theta+1)}\ e^{-\theta x}, \quad x > 0.$$

A neutrosophic random variable $X_N = [X_L, X_U]$ follows a neutrosophic ZLD (NZLD) with neutrosophic scale parameter $\theta_N \in [\theta_L,\, \theta_U]$, if its neutrosophic pdf (npdf) is

$$f_N(x_N,\theta_N) = \frac{\theta_N(1+2\theta_N+\theta_N x_N)}{2(\theta_N+1)}\ e^{-\theta_N x_N}, \tag{1}$$

where X_L and X_U (or θ_L and θ_U) represent the lower bound and upper bound of X_N (or θ_N). $f_N(x_N,\theta_N)$ given in (1) defines a valid neutrosophic pdf on $(0,+\infty)$, since

$$\begin{aligned}
\int_0^{+\infty} f_N(x_N;\theta_N)dx_N &= \int_0^{+\infty} \frac{\theta_N(1+2\theta_N+\theta_N x_N)}{2(\theta_N+1)}\ e^{-\theta_N x_N} dx_N \\
&= \frac{\theta_N}{2(\theta_N+1)} \left(\int_0^{+\infty} (1+2\theta_N)\ e^{-\theta_N x_N} dx_N + \int_0^{+\infty} \theta_N x_N\ e^{-\theta_N x_N} dx_N\right) \\
&= \frac{\theta_N}{2(\theta_N+1)} \left(\frac{2\theta_N+1}{\theta_N} + \frac{1}{\theta_N}\right) \\
&= 1.
\end{aligned}$$

The corresponding neutrosophic survival function of $f_N(x_N,\theta_N)$ is given by

$$S_N(x_N,\theta_N) = 1 - F_N(x_N,\theta_N) = \left(1 + \frac{\theta_N x_N}{2(\theta_N+1)}\right) e^{-\theta_N x_N}, \quad x_N > 0. \tag{2}$$

while the neutrosophic hazard function (NHF) of the NZLD is

$$h_N(x_N;\theta_N) = \frac{(1+2\theta_N+\theta_N x_N)\theta_N}{2+2\theta_N+\theta_N x_N}, \quad x_N > 0. \tag{3}$$

The plot of the npdf of the NZLD is shown in Figure 1, while the plots of the ncdf and NHF of the NZLD are presented in Figures 2 and 3, respectively. In Figure 1, $\theta_N \in [0.5, 1.0], [0.25, 0.40], [0.55, 2.0]$ and $[2, 3.5]$ from left panel to right. Figure 2, $\theta_N \in [0.5, 1.0], [0.5, 2.0], [1.5, 3.0]$ and $[0.1, 0.35]$ from left panel to right. Figure 3, $\theta_N \in [0.05, 0.1], [1.5, 2.0], [0.35, 0.45]$ and $[2.25, 3.5]$ from left panel to right. In all figures, the black curves represent the function values at the lower and upper bounds of the neutrosophic scale parameter θ_N. The colored regions illustrate how the function values vary as the parameter shifts across its indeterminacy interval.

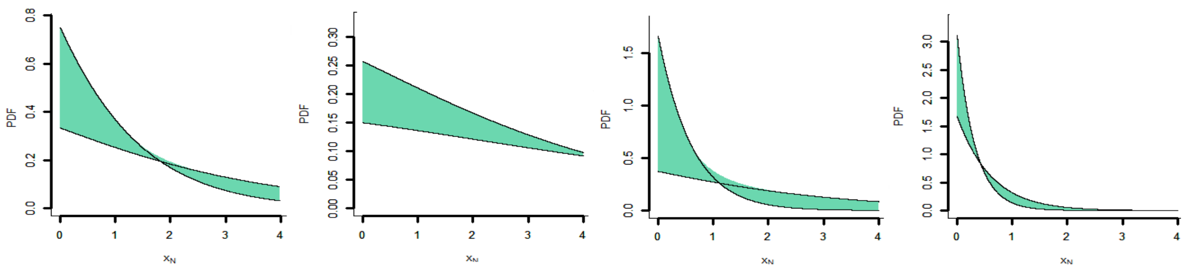

Figure 1: The plot of npdf of the NZLD in (1).

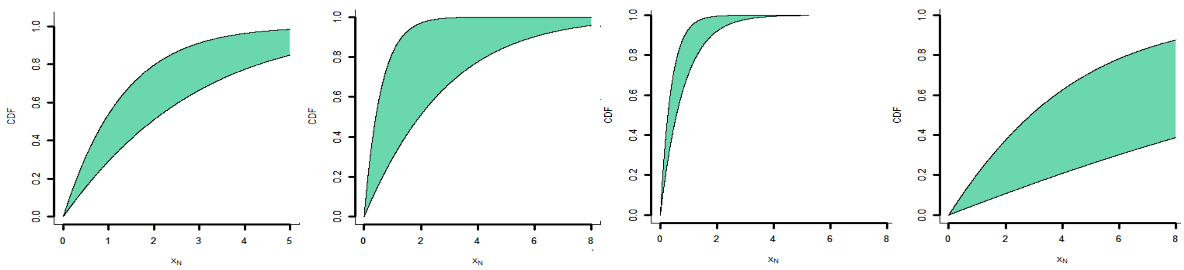

Figure 2: The plot of $F_N(x_N, \theta_N)$ in (2).

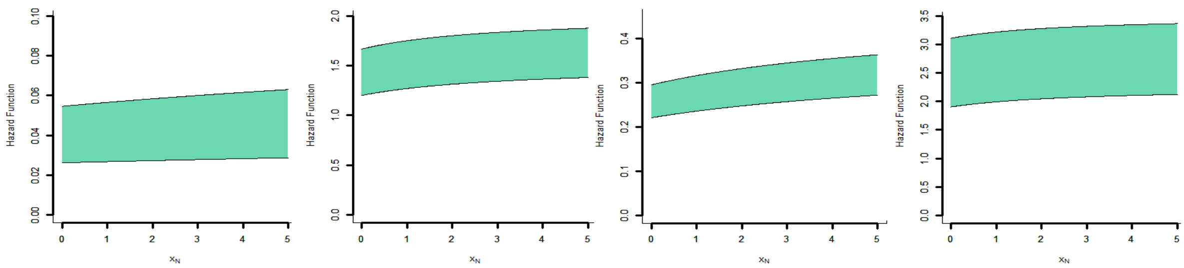

Figure 3: the plot of NHF of the NZLD in (3).

3 Statistical properties of the NZLD

Suppose X has npdf $f_N(x_N,\theta_N)$ as given in (1). Then, it is easy to show that the r-th raw moment of X_N is

$$\mu_r' = E(X_N^r;\theta_N) = \frac{(r+2+2\theta_N)\Gamma(r+1)}{2\theta_N^r(1+\theta_N))},$$

where $\Gamma(.)$ is the complete gamma function. Hence, we the first four raw moments are

$$\mu_1' = \frac{(3+2\theta_N)}{2\theta_N(1+\theta_N)}, \mu_2' = \frac{2(2+\theta_N)}{\theta_N^2(1+\theta_N)}, \mu_3' = \frac{3(5+2\theta_N)}{\theta_N^3(1+\theta_N)}, \text{ and } ,\mu_4' = \frac{24(3+\theta_N)}{\theta_N^4(1+\theta_N)}.$$

Using the r-th raw moment, we can determine the p-th central moments of X_N as follows

$$\mu_p = E(X_N-\mu_1')^p = \sum_{\ell=0}^{p}(-1)^\ell\binom{p}{\ell}(\mu_1')^{p-\ell}\mu_{p-\ell}'.$$

We used μ_r', $r=1,\ldots,4$, to obtain the variance, coefficient of variation (CV), skewness (γ_1), and kurtosis (γ_2). They are given as follows:

$$Var(X_N) = \frac{(4\theta_N^2+12\theta_N+7)}{4\theta_N^2(\theta_N+1)^2},$$

$$CV = \frac{\sqrt{Var(X_N)}}{\mu_1'} = \frac{\sqrt{4\theta_N^2+12\theta_N+7}}{3+2\theta_N},$$

$$\gamma_1 = \frac{\mu_3'-3\mu_2'\mu_1'+2(\mu_1')^3}{(\mu_2'-(\mu_1')^2)^{3/2}} = \frac{2(8\theta_N^3+36\theta_N^2+42\theta_N+15)}{(4\theta_N^2+12\theta_N+7)^{3/2}},$$

and

$$\gamma_2 = \frac{\mu_4'-4\mu_3'\mu_1'+6\mu_2'(\mu_1')^2-3(\mu_1')^4}{(\mu_2'-(\mu_1')^2)^2} = \frac{144\theta_N^4+864\theta_N^3+1608\theta_N^2+1224\theta_N+333}{(4\theta_N^2+12\theta_N+7)^2}.$$

For $\theta_N>0$, the expressions for Var(X_N), CV, γ_1, and γ_2 are differentiable with respect to θ_N. We have

$$\frac{d\mu_1'}{d\theta_N} = -\frac{2\theta_N^2+6\theta_N+3}{2\theta_N^2(1+\theta_N)^2},$$

$$\frac{dVar(X_N)}{d\theta_N} = -\frac{4\theta_N^3+18\theta_N^2+20\theta_N+7}{2\theta_N^3(\theta_N+1)^3},$$

$$\frac{dCV}{d\theta_N} = -\frac{10\theta_N+9}{(3\theta_N+2)^2\sqrt{4\theta_N^2+12\theta_N+7}},$$

$$\frac{d\gamma_1}{d\theta_N} = \frac{96\theta_N^2+144\theta_N+48)}{(4\theta_N^2+12\theta_N+7)^{5/2}},$$

and

$$\frac{d\gamma_2}{d\theta_N} = \frac{192(8\theta_N^3+18\theta_N^2+13\theta_N+3)}{(4\theta_N^2+12\theta_N+7)^3}.$$

Since θ_N is positive, $\frac{d\mu_1'}{d\theta_N}, \frac{dCV}{d\theta_N}$ and $\frac{dVar(X_N)}{d\theta_N}$ are strictly negative while $\frac{d\gamma_1}{d\theta_N}$ and $\frac{d\gamma_2}{d\theta_N}$ are strictly positive. These imply that Var(X_N) and CV are decreasing in θ_N while γ_1, and γ_2 are increasing. Table 1 contains

Table 1: Mean, variance, coefficient of variation, skewness and kurtosis of the NZLD.

θ_N	μ_1'	$Var(X_N)$	CV	γ_1	γ_2
[0.10,0.25]	[5.6,14.545]	[26.24,170.248]	[0.897,0.915]	[1.655,1.699]	[6.957,7.173]
[0.25,0.50]	[2.667,5.600]	[6.222,26.24]	[0.915,0.935]	[1.699,1.756]	[7.173,7.469]
[0.80,1.00]	[1.250,1.597]	[1.438,2.310]	[0.952,0.959]	[1.806,1.831]	[7.743,7.888]
[1.25,1.75]	[0.675,0.978]	[0.434,0.893]	[0.966,0.976]	[1.856,1.892]	[8.038,8.259]
[2.00,2.50]	[0.457,0.583]	[0.202,0.326]	[0.979,0.984]	[1.906,1.925]	[8.343,8.472]
[2.50,3.50]	[0.375,0.457]	[0.137,0.202]	[0.984,0.988]	[1.925,1.94]	[8.472,8.567]

the values of variance, CV, skewness (γ_1), and kurtosis (γ_2) for some values selected of θ_N under the neutrosophic framework. Table 1 shows that the mean, variance, and coefficient of variation decrease as θ_N increases, whereas the skewness and kurtosis increase. This behavior is also evident from their formulas as θ_N increases.

4 MLE of neutrosophic parameter

Let $x_1, x_2, \ldots, x_n$ be observed values of $X_1, X_2, \ldots, X_n$, where $X_1, \ldots, X_n$ are independent random variables following the ZLindley distribution with unknown parameter θ. Then, the log-likelihood function is

$$\ln L(\theta; x_1, \ldots, x_n) = n\ln\theta - n\ln(2) - n\ln(\theta+1) + \sum_{i=1}^{n}\ln(1+2\theta+\theta x_i) - \theta\sum_{i=1}^{n}x_i.$$

The MLE of θ is obtained by solving the non-linear equation $\frac{\partial}{\partial\theta}\ln L(\theta; x_1, \ldots, x_n) = 0$, where

$$\frac{\partial}{\partial\theta}\ln L(\theta; x_1, \ldots, x_n) = \frac{n}{\theta(\theta+1)} + \sum_{i=1}^{n}\frac{2+x_i}{1+2\theta+\theta x_i} - \sum_{i=1}^{n}x_i. \tag{4}$$

We extend the maximum likelihood framework to account for the uncertainty of the data by defining each observation as an interval $x_{Ni} = [x_{Li}, x_{Ui}]$ for $i = 1, \ldots, n$. Consequently, the neutrosophic log-likelihood function is

$$\ln L(\theta_N; x_{N1}, \ldots, x_{Nn}) = n\ln(\theta_N) - n\ln(2) - n\ln(\theta_N+1) + \sum_{i=1}^{n}\ln\big(1+2\theta_N+\theta_N x_{Ni}\big) - \theta_N\sum_{i=1}^{n}x_{Ni}.$$

In neutrosophic statistics, the likelihood function is optimized over the interval space of the data to account for indeterminacy. Consequently, we use a two-step interval maximization to obtain the neutrosophic MLE (NMLE):

- The first step is to define the indeterminate space $S_N = \prod_{i=1}^{n} x_{Ni}$ which is the space of all possible observations, where each observation interval x_{Ni} adds to the total uncertainty of the parameter.
- The second step is to determine the MLE of parameters for each $\mathbf{x} = (x_{N1}, \ldots, x_{Nn}) \in S_N$ using equation (4). The minimum and maximum estimates for each of the possible values of $\mathbf{x}$ are then calculated to get the NMLE.

For a given interval observation $\mathbf{x} \in S_N$. Let $\Theta(\mathbf{x}) = \arg\max_{\theta_N}\ln L(\theta, x_{N1}, \ldots, x_{Nn})$ represents the set of parameter values that maximize the likelihood function. Therefore, the following formula can be used to calculate the NMLE:

$$\hat{\theta}_N = \left(\min_{x\in S_N}\{\theta : \theta \in \Theta(x)\},\ \max_{x\in S_N}\{\theta : \theta \in \Theta(x)\}\right). \tag{5}$$

A large number of data vectors are generated within the admissible region S_N to carry out the nonlinear optimization in (4). Each vector is created by independently sampling values from the observed intervals. For each generated vector, the corresponding MLE is recorded after maximizing the likelihood function with respect to the parameter. Finally, the extreme values of these estimates across all evaluated instances are extracted to obtain $\hat{\theta}_N$.

5 Simulation study

The performance of NMLE in the NZLD is assessed through a simulation study. The NZLD provides interval-valued estimates that account for both inherent uncertainty and randomness, in contrast to the point estimates provided by classical statistics. A neutrosophic variable is defined as $X_N = X + I$, where X is the determined part, and I is the uncertain part. Within this framework, I can be represented as a uniform distribution $I \sim U(-\varepsilon, \varepsilon)$ to represent variable uncertainty, or it can be a constant value that captures fixed uncertainty. This method improves the accuracy of real-world estimations by allowing the model to incorporate both static and dynamic uncertainties. In our simulation study, we generate $N = 1000$ random samples of sizes $n = 10, 20, 50, 200$, and 300 from NZLD with $\theta = 1.25$ and $\theta = 2$. Additive uniform noise $I \sim U(-\varepsilon, \varepsilon)$ was used to introduce uncertainty for each observation x_N, with $\varepsilon = 0.01, 0.05, 0.1$ and 0.2. This approach reflects the model's adaptability to uncertain data and captures variability. Using (5), we compute NMLE to evaluate the performance. To this end, we calculate the neutrosophic mean squared error (NMSE), neutrosophic average estimates (NAE) and neutrosophic average biases (NAB):

$$\mathrm{NAE} = \frac{1}{1000}\sum_{i=1}^{1000}\theta_{N_i}, \;\; \mathrm{NAB} = \frac{1}{1000}\sum_{i=1}^{1000}\left(\theta_{N_i} - \theta_N\right), \;\text{ and }\; \mathrm{MSE} = \frac{1}{1000}\sum_{i=1}^{1000}\left(\theta_{N_i} - \theta_N\right)^2.$$

Tables 2 and 3 confirm that the estimator of the parameter of the NZLD works consistently with more

Table 2: NAE, NAB and NMSE for θ =1.25.

ε	n	NAE	NAB	NMSE
0.01	10	[1.3424,1.3498]	[0.0924,0.0998]	[0.4376,0.4454]
	50	[1.2575,1.2634]	[0.0075,0.0134]	[0.1537,0.1557]
	100	[1.2536,1.2594]	[0.0036,0.0094]	[0.1053,0.1066]
	300	[1.2522,1.258]	[0.0022,0.0080]	[0.0608,0.0618]
0.05	10	[1.328,1.3653]	[0.0780,0.1153]	[0.4225,0.4620]
	50	[1.2459,1.2754]	[-0.0041,0.0254]	[0.1505,0.1604]
	100	[1.2422,1.2712]	[-0.0078,0.0212]	[0.1035,0.1104]
	300	[1.2409,1.2697]	[-0.0091,0.0197]	[0.0603,0.0656]
0.1	10	[1.3107,1.3853]	[0.0607,0.1353]	[0.4055,0.4850]
	50	[1.2317,1.2908]	[-0.0183,0.0408]	[0.1480,0.1678]
	100	[1.2282,1.2863]	[-0.0218,0.0363]	[0.1032,0.1170]
	300	[1.227,1.2847]	[-0.023,0.0347]	[0.0626,0.0730]
0.2	10	[1.2779,1.4283]	[0.0279,0.1783]	[0.3771,0.5411]
	50	[1.2044,1.3229]	[-0.0456,0.0729]	[0.1474,0.1869]
	100	[1.2011,1.3175]	[-0.0489,0.0675]	[0.1082,0.1354]
	300	[1.2001,1.3158]	[-0.0499,0.0658]	[0.0748,0.0945]

data and effectively reduces estimation error by demonstrating that NMSE decreases with larger sample sizes. Moreover, the parameter intervals widen as ε increases, indicating the flexibility of the model in the face of increased uncertainty. This procedure provides more accurate estimates in the face of data indeterminacy and is consistent with the behavior predicted by the interval neutrosophic framework.

Table 3: NAE, NAB and NMSE for θ =2.

ε	n	NAE	NAB	NMSE
0.01	10	[2.1509,2.1714]	[0.1509,0.1714]	[0.7165,0.7389]
	50	[2.0096,2.0257]	[0.0096,0.0257]	[0.2499,0.2554]
	100	[2.0031,2.0189]	[0.0031,0.0189]	[0.1711,0.1749]
	300	[2.0006,2.0163]	[0.0006,0.0163]	[0.0987,0.1017]
0.05	10	[2.1121,2.2147]	[0.1121,0.2147]	[0.6758,0.7881]
	50	[1.9783,2.0589]	[-0.0217,0.0589]	[0.2423,0.2700]
	100	[1.9722,2.0513]	[-0.0278,0.0513]	[0.1680,0.1872]
	300	[1.9700,2.0486]	[-0.0300,0.0486]	[0.1002,0.1148]
0.1	10	[2.0663,2.2727]	[0.0663,0.2727]	[0.6332,0.8629]
	50	[1.9407,2.1021]	[-0.0593,0.1021]	[0.2393,0.2945]
	100	[1.9351,2.0935]	[-0.0649,0.0935]	[0.1722,0.2103]
	300	[1.9331,2.0905]	[-0.0669,0.0905]	[0.1137,0.1416]
0.2	10	[1.9825,2.4051]	[-0.0175,0.4051]	[0.5709,1.0877]
	50	[1.8700,2.1949]	[-0.1300,0.1949]	[0.2512,0.3626]
	100	[1.8651,2.1837]	[-0.1349,0.1837]	[0.2006,0.2771]
	300	[1.8635,2.1801]	[-0.1365,0.1801]	[0.1611,0.2164]

6 Applications

In this section, two data sets are used to illustrate the applicability of the proposed distribution.

6.1 Pseudo neutrosophic data

In order to prove the potentiality of the NZLD, we give an application using a well-known real data set. The dataset is given by Ghitany el al. (2008), and represents the waiting times (in minutes) before service of 100 bank customers. The dataset is:
0.8, 0.8, 1.3, 1.5, 1.8, 1.9, 1.9, 2.1, 2.6, 2.7, 2.9, 3.1, 3.2, 3.3, 3.5, 3.6, 4.0, 4.1, 4.2, 4.2, 4.3, 4.3, 4.4, 4.4, 4.6, 4.7, 4.7, 4.8, 4.9, 4.9, 5.0, 5.3, 5.5, 5.7, 5.7 ,6.1, 6.2, 6.2, 6.2, 6.3, 6.7, 6.9, 7.1, 7.1, 7.1, 7.1, 7.4, 7.6, 7.7, 8.0, 8.2, 8.6, 8.6, 8.6, 8.8, 8.8, 8.9, 8.9, 9.5, 9.6, 9.7, 9.8, 10.7, 10.9, 11.0, 11.0, 11.1, 11.2, 11.2, 11.5, 11.9, 12.4, 12.5, 12.9, 13.0, 13.1, 13.3, 13.6, 13.7, 13.9, 14.1, 15.4, 15.4, 17.3, 17.3, 18.1, 18.2, 18.4, 18.9, 19.0, 19.9, 20.6, 21.3, 21.4, 21.9, 23.0, 27.0, 31.6, 33.1 and 38.5.

We use the parameter ε to introduce different levels of indeterminacy in order to extend the analysis into the neutrosophic framework. The exact dataset is used to build neutrosophic datasets, in which a certain amount of uncertainty is introduced by each ε value. To represent varying degrees of indeterminacy, we use $\varepsilon = 0, 0.001, 0.01, 0.1$ and 1 and take into consideration $X_N = X + \varepsilon$. The data is accurate and represents a classical situation when $\varepsilon = 0$. The dataset shows increasing degrees of uncertainty as ε increases, which is consistent with the neutrosophic notion. We estimate the parameter θ_N of the NZLD model across the various ε-values using the transformed datasets. Table 3 displays these estimates as well as the corresponding Bayesian Information Criterion (BIC) and Akaike's Information Criterion (AIC).

To extend the analysis into the neutrosophic framework, we introduce varying levels of indeterminacy using the parameter ε. The precise dataset serves as the basis for creating neutrosophic datasets, where each ε value introduces a specific degree of uncertainty. We take $X_N = X + \varepsilon$ and set $\varepsilon = 0, 0.001, 0.01, 0.1$, and 0.75 to show different levels of indeterminacy. When $\varepsilon = 0$, the dataset remains precise, reflecting a classical scenario. As ε increases, the dataset captures increasing levels of uncertainty, aligning with the neutrosophic concept.
Using the transformed datasets, we estimate the parameter θ of the NZLD model for different ε-values.

This estimate, the Akaike's Information Criterion (AIC), and the Bayesian Information Criterion (BIC) are presented in table 4. The AIC and BIC interval values reveal how well the neutrosophic model fits the data at various levels of uncertainty. The neutrosophic model approaches the behavior of classical statistical models, which rely on exact data, as the degree of uncertainty decreases (i.e., smaller ε-values). This specifically corresponds to the ZLD when $\varepsilon = 0$. However, the model captures more uncertainty with larger ε-values, indicating the adaptability and resilience of the neutrosophic approach. Since θ_N is unknown, we evaluate the model's fit using the modified Kolmogorov-Smirnov (KS*) test.

Table 4: NAE, NAB and NMSE for θ =1.25.

ε	θ_N	Log-likelihood	AIC	BIC	KS*
0	0.1481	-325.2583	652.5167	655.1219	0.5076
0.001	[0.14805,0.14808]	[-325.267,-325.250]	[652.499,652.534]	[655.104,655.140]	[0.5075,0.5078]
0.01	[0.1479,0.1482]	[-325.346,-325.171]	[652.342,652.692]	[654.947,655.297]	[0.506,0.509]
0.1	[0.1468,0.1494]	[-326.131,-324.379]	[650.759,654.263]	[653.364,656.866]	[0.491,0.523]
0.75	[0.1387,0.1587]	[-331.641,-318.495]	[638.989,665.282]	[641.595,667.887]	[0.599,0.610]

6.2 Neutrosophic failure time dataset

This dataset, provided by Linhart and Zucchini (1986, p. 69), represents the failure times of the air conditioning system of an airplane. The corresponding neutrosophic data are:
23, 261, 87, 7, [121,125], 14, 62, 47, 225, [71,74], 246, 21, 42, 20, 5, 12, 120, [11,13], 3, 14, 71, 11, 14, 11, 16, 90, 1, 16, [52,55] and 95.

We compare the performance of the NZLD with the neutrosophic Lindley distribution (NLD), proposed by Bashir el al. (2025). From Table 5, the NZLD model provides the lowest AIC, BIC, CAIC, HQIC, and KS* values, indicating superior performance compared to the NLD.

Table 5: Comparison between the NZLD and NLD.

Models	NZLindley	NLindley
NMLE of θ_N	[0.0238,0.0240]	[0.0328,0.0330]
Log-L	[-154.508,-154.338]	[-161.8429,-161.5933]
AIC	[310.676,311.017]	[325.1866,325.6859]
BIC	[312.077,312.418]	[326.5878,327.0871]
CAIC	[313.077,313.418]	[327.5878,328.0871]
HQIC	[311.124,311.465]	[325.6348,326.1341]
KS*	[0.0199,0.0234]	[0.0240,0.0269]
p-value	[0.9670,0.9820]	[0.9527,0.9679]

7 Conclusion

Within a neutrosophic framework, we introduced an extension of the ZLindley distribution, termed the neutrosophic ZLindley distribution (NZLD). Several properties of the proposed model were derived, and the model parameter was estimated using NMLE. A simulation study demonstrated that this estimator performs best under both bias and MSE criteria. Additionally, a well-known real dataset was analyzed, and the results indicate that the proposed model provides a superior fit compared to the NLD.

References

Alduais, F. S., and Khan, Z. (2025). Development of neutrosophic gamma distribution for modeling neonatal mortality data. *Neutrosophic Sets and Systems*, **79**, 48.

Alhabib, R., Ranna, M. M., Farah, H. and Salama, A. A. (2018). Some neutrosophic probability distributions. *Neutrosophic Sets and Systems*, **22**, 30–38.

Alhasan, K. F. H. and Smarandache, F. (2019). Neutrosophic Weibull distribution and neutrosophic family Weibull distribution. *Neutrosophic Sets and Systems*, **28**, 191–199.

Bashir, S., Masood, B., Shehzadi, I. and Aslam, M. (2024). Neutrosophic log-Logistic distribution: a novel approach for modeling uncertain survival and reliability data. *REVSTAT-Statistical Journal*. Accepted-November 2024. https://revstat.ine.pt/index.php/REVSTAT/article/view/789.

Bashir, S., Masood, B., Shehzadi, I., Al-Husseini, Z. and Aslam, M. (2025). Neutrosophic Lindley distribution: simulation, application, and comparative study. *Contemporary Mathematics*, **2025**, 551–564.

Ghitany, M. E, Atieh, B. and Nadarajah, S. (2008). Lindley distribution and its application. *Mathematics and Computers in Simulation*, **78**, 493–506.

Khan Sherwani, R.A., Naeem, M., Aslam, M., and Raza, M. A. (2021). Neutrosophic beta distribution with properties and applications. *Neutrosophic Sets and Systems*, **41**, 12–19.

Linhart, H. and Zucchini, W. (1986). *Model Selection*. John Wiley, New York.

Razmkhah, M, Arashi, A., Bekker, A. and Marques, F. J. (2026). Neutrosophic Birnbaum-Saunders distribution with applications. *Applied Mathematical Modelling* , **149**, 116287.

Saaidia, N., Belhamra, T. and Zeghdoudi, H. (2024). On ZLindley distribution: statistical properties and applications. *Studies in Engineering and Exact Sciences*, **5**, 3078–3097.

Smarandache, F. (2014). *Introduction to Neutrosophic Statistics*. Sitech and Education Publishing, Columbus, Ohio.

Stochastic Models in Probability and Statistics
Vol. 2, No. 2, 2025, pp. 221–231. Research Paper

Probabilistic metric structures on birth-death population models

T. K. Murali†*, P. K. Pramod‡, P. B. Vinod Kumar§

†*Department of Mathematics, KKTM Govt. College, Pullut, Kerala, India*
†*Department of Mathematics, Maharaja's College, Ernakulam, Kerala, India*
‡*Department of Mathematics, Panambilly Memorial Govt. College, Chalakkudy, Kerala, India*
‡*Department of Mathematics, Maharaja's College, Ernakulam, Kerala, India*
§*Department of Basic Science and Humanities, Muthoot Institute of Technology and Science(MITS), Kochi, Kerala, India*

Email(s): muralitk.maths@gmail.com, pramodnair.2007@gmail.com,vinodkumarpb@mgits.ac.in

Abstract. In this paper, we introduce a probabilistic metric space structure for a population genetics model that incorporates mutation effects. Several properties of this structure are investigated in relation to the mutation process.

Keywords: Chapman-Kolmogorov equation, Markov Chains, Population genetics, Probabilistic metric spaces.

1 Introduction

A metric space is a nonempty set in which a non-negative value is assigned as the distance between each pair of points. In a probabilistic metric space, instead of assigning a deterministic non-negative number, a distance distribution function is associated with each pair of points. This function can sometimes be interpreted as the probability that the distance between the points does not exceed a given value.
Birth-death processes are stationary Markov processes in which only two types of transitions are possible, namely birth and death. It is typically assumed that, within a sufficiently small time interval, at most one birth or one death can occur. Using the probabilities of a single birth or death, transition probabilities are defined and a transition probability matrix is constructed. When the state space of a birth-death process is a metric space, a probabilistic metric space structure can be induced by combining the underlying metric with the transition probabilities.

Examining changes in the genetic composition of a population as a result of various evolutionary processes, including selection, mutation, migration, and random genetic drift, is one of the central objectives of theoretical population genetics. Consider a population of haploid individuals with a constant fixed size N. We focus on a single locus that can carry two types of alleles, A and B. Different alleles may lead to differences in fitness, meaning that some individuals are more strongly selected than others. If mutations can occur in both directions, the population will always remain polymorphic, and in this case

*Corresponding author

Received: 17 June 2025 / Accepted: 20 December 2025
DOI: 10.22067/smps.2025.94103.1048

 https://smps.um.ac.ir

the stationary distribution becomes the primary object of interest. To study these phenomena statistically, a suitable mathematical model is required. When attention is restricted to individuals of type A, mutation from A to B may be interpreted as a death event, while mutation from B to A corresponds to a birth event. Birth and death rates can be defined mathematically, and the mutation process can thus be modeled as a birth-death process. Using the transition probabilities associated with this process, the population consisting of two types of individuals can be endowed with a probabilistic metric space structure, allowing further properties of the mutation dynamics to be explored from this perspective. This construction is carried out in three different ways with slight modifications. In the first case, only the possible population sizes at different time points are considered as distinct states of the process. In the second and third cases, individual elements are treated as distinct, and the power set of the population is taken as the state space. In each case, a probabilistic metric space is constructed by taking into account the corresponding variations.

In Schweizer and Sklar (2005), probabilistic metric spaces are introduced and studied in detail. Markov chains and their fundamental properties are discussed in Bhat and Miller (1984), while applications of Markov chains in genetics are presented in Karlin (1975). The probabilistic metric space structure induced by stationary Markov processes is investigated in Marcus (1977). A similar idea is developed in Moynihan (1976), where stationary Markov processes with metric state spaces are considered. In Murali et al. (2022), three different cases of probabilistic metric spaces induced by birth-death processes are examined, along with related results.

In the present paper, we discuss simple examples of probabilistic metric spaces arising from genetic models and explore the applicability of the results in Murali et al. (2022) to the modeling of mutation processes in population genetics.

2 Preliminaries

2.1 Markov chains

In this paper, we employ Markov chains and their transition probabilities as follows. The number of members in the population at each stage is taken as the state space. Thus, the state space is $S=\{1,2,\ldots\}$. Let $\{X_n\}_{n\in\mathbb{N}\cup\{0\}}$ denote a Markov chain defined on this state space. The probability distribution of X_{n+1} depends only on the value of X_n at time n and not on the previous states. Formally,

$$P\big(X_{n+1}=a_{n+1}\mid X_0=a_0,X_1=a_1,\ldots,X_n=a_n\big)=P\big(X_{n+1}=a_{n+1}\mid X_n=a_n\big).$$

We consider only time-homogeneous Markov chains. This property is expressed mathematically as

$$P(X_{n+1}=a_j\mid X_n=a_i)=P(X_{m+1}=a_j\mid X_m=a_i)\quad\text{for all } n,m.$$

Let $P_{ij}^{(n)}=P(X_{m+n}=a_j\mid X_m=a_i)$ denote the n-step transition probabilities. By the Chapman-Kolmogorov equations, we have $P_{ij}^{(r+s)}=\sum_{k\in S}P_{ik}^{(r)}P_{kj}^{(s)}$.

2.2 Metric spaces to probabilistic metric spaces

Definition 1. *A* metric space *is an ordered pair (S,d), where S is a nonempty set and $d\colon S\times S\to\mathbb{R}^+\cup\{0\}$ is a function satisfying the following properties for all $a,b,c\in S$:*

(i) $d(a,b)\geq 0$;

(ii) $d(a,b)=0$ *if and only if* $a=b$;

(iii) $d(a,b)=d(b,a)$;

(iv) $d(a,c) \leq d(a,b) + d(b,c)$.

The value $d(a,b)$ is interpreted as the distance between the points a and b. Thus, in a metric space, a deterministic nonnegative distance is assigned to each pair of points. In contrast, in a probabilistic metric space, a distance distribution function is assigned to each pair of points. This function may be interpreted as describing the probability that the distance between the two points does not exceed a given value.

Definition 2. *A non decreasing function F defined on R^+ which satisfies $F(0) = 0$, $F(\infty) = 1$ and left continuous on $(0,\infty)$ is a distance distribution function.*

The collection of all distance distribution functions is denoted by Δ^+. For $a \geq 0$, the function $\varepsilon_a(x)$ is defined by

$$\varepsilon_a(x) = \begin{cases} 1, & x > a, \\ 0, & x \leq a. \end{cases} \tag{1}$$

Definition 3. *A binary operation on Δ^+ is called a* triangle function *if it is commutative, associative, non-decreasing in each argument, and has ε_0 as the identity element.*

Equivalently, a map $\tau \colon \Delta^+ \times \Delta^+ \to \Delta^+$ is called a triangle function if it satisfies the following conditions:

1. *τ is commutative, that is, $\tau(F,G) = \tau(G,F), \quad \forall F,G \in \Delta^+$;*
2. *τ is associative, that is, $\tau\big(\tau(F,G),H\big) = \tau\big(F,\tau(G,H)\big), \quad \forall F,G,H \in \Delta^+$;*
3. *τ is non-decreasing in each argument, i.e., $F_1 \leq F_2 \Rightarrow \tau(F_1,G) \leq \tau(F_2,G), \quad \forall G \in \Delta^+$, and $G_1 \leq G_2 \Rightarrow \tau(F,G_1) \leq \tau(F,G_2), \quad \forall F \in \Delta^+$;*
4. *ε_0 is the identity element, that is, $\tau(F,\varepsilon_0) = F, \quad \forall F \in \Delta^+$.*

The product operation is a triangle function, defined by $(\tau(F,G))(x) = F(x)\,G(x)$. Also, the triangle function generated by the Łukasiewicz t-norm $T_m(a,b) = \max\{a+b-1, 0\}$ is denoted by τ_{T_m} and is defined as

$$(\tau_{T_m}(F,G))(x) = \sup_{0 \leq u \leq x} T_m\big(F(u), G(x-u)\big).$$

Definition 4. *Let $(S,\mathscr{F},\tau)$ be a triple, where S is a nonempty set, $\mathscr{F} : S \times S \to \Delta^+$ is a function, and τ is a triangle function. The triple $(S,\mathscr{F},\tau)$ is called a* probabilistic metric space *if, for all $p,q,r \in S$, the following conditions hold:*

(i) $\mathscr{F}(p,p) = \varepsilon_0$;

(ii) $\mathscr{F}(p,q) \neq \varepsilon_0$ *whenever* $p \neq q$;

(iii) $\mathscr{F}(p,q) = \mathscr{F}(q,p)$;

(iv) $\mathscr{F}(p,r) \geq \tau\big(\mathscr{F}(p,q), \mathscr{F}(q,r)\big)$.

If only conditions (iii) and (iv) are satisfied, then $(S,\mathscr{F},\tau)$ is called a pseudo probabilistic metric space.

The above definition is well defined. Indeed, since $\tau \colon \Delta^+ \times \Delta^+ \to \Delta^+$, for any $p,q,r \in S$ we have $\tau\big(\mathscr{F}(p,q), \mathscr{F}(q,r)\big) \in \Delta^+$. Moreover, Δ^+ is a partially ordered set, where for $A,B \in \Delta^+$, $A \leq B$ if and only if $A(x) \leq B(x)$ for all $x \geq 0$. Consequently, the inequality $\mathscr{F}(p,r) \geq \tau\big(\mathscr{F}(p,q), \mathscr{F}(q,r)\big)$ is meaningful and mathematically consistent.

Example 1. (Metric spaces as probabilistic metric spaces) Let (S,d) be a metric space. Take $F(p,q) = \varepsilon_{d(p,q)}$ where ε_d is given in (1). Also put $\tau(\varepsilon_a, \varepsilon_b) = \varepsilon_{a+b}$ then for any $p,q,r \in S$,

$$\tau(F(p,q), F(q,r)) = \varepsilon_{d(p,q)+d(q,r)}.$$

Hence, every metric space yields a Probabilistic metric space.

Example 2. (Genetic model) Consider the state space, $S = \{A, a\}$(two alleles at a single locus). We model mutation as a continuous- time Markov chain (CTMC) with constant rates:

$$A \xrightarrow{\mu} a, \ a \xrightarrow{\nu} A,$$

where $\mu, \nu > 0$ are mutation rates (per unit time). Let $(X_t)_{t\geq 0}$ be this CTMC and P_x denote probability with $X_0 = x$. Then, for $x, y \in S$, define

$$F_{xy}(t) = P_x\big(\exists\ s \leq t \text{ with } X_s = y\big),$$

as hitting distribution function. We will compute $F_{Aa}(t)$ and $F_{aA}(t)$ explicitly. Starting at A, the chain jumps away from A after an exponential time with parameter μ (the only outgoing rate). Thus, the time $T_{A\to a}$ to first hit a (which is the first jump time) is $\mathrm{Exp}(\mu)$. Therefore,

$$F_{Aa}(t) = P(T_{A\to a} \leq t) = 1 - e^{-\mu t}.$$

Similarly, starting at a, we have

$$F_{aA}(t) = 1 - e^{-\nu t}.$$

By Definition 2, it readily follows that F_{Aa} and F_{aA} belong to Δ^+.

Remark 1. *$F_{xy} \neq \varepsilon_0$ for $x \neq y$ (since $F_{xy}(t) > 0$ for $t > 0$), and $F_{xx} = 1$ for all $t \geq 0$ if we define "hit itself by time t" trivially. To match the condition $F(p,p) = \varepsilon_0$, we instead define $F_{xx}(t)$ as the degenerate distribution: the time to hit itself is 0 with probability 1. Formally, take $F_{xy} = \varepsilon_0(t) = \mathbf{1}_{t\geq 0}$.*

• **Triangle inequality (probabilistic version).** Pick three states $p,q,r \in S$. There are only two nontrivial distinct-state choices; we will show the triangle-type inequality

$$F_{pr}(t) \geq \int_0^t dF_{pq}(S) F_{qr}(t-S). \tag{2}$$

This inequality is the probabilistic analogue of the triangle inequality and is obtained from the Markov property by conditioning on the first hitting time of q.

To Prove, (2), Let $T_{p\to q} := \inf\{s \geq 0 : X_s = q\}$ be the first hitting time of q starting from p. Then

$$\{\text{hit } r \text{ by time } t\} \supseteq \{T_{p\to q} \leq t \text{ and then hit } r \text{ within } t - T_{p\to q}\}.$$

Therefore, we have

$$\begin{aligned} F_{pr}(t) &= P_p(\exists\ s \leq t : X_s = r) \\ &\geq P_p(T_{p\to q} \leq t, \exists\ u \leq t - T_{p\to q} : X_{T_{p\to q}+u} = r). \end{aligned}$$

By the strong Markov property, conditional on $T_{p\to q} = s$ and $X_s = q$, the future evolution $\{X_{s+u} : u \geq 0\}$ has the same law as the chain started at q. Hence

$$\begin{aligned} P_p(T_{p\to q} \leq t, \exists\ u \leq t - T_{p\to q} : X_{T_{p\to q}+u} = r) &= \int_0^t P_p(T_{p\to q} \in ds)\ P_q(\exists\ u \leq t-s : X_u = r) \\ &= \int_0^t dF_{pq}(s) F_{qr}(t-s). \end{aligned}$$

Interpreting (2) as a triangle function inequality

Define a triangle function τ on Δ^+ by

$$(\tau(F,G))(t) := \int_0^t dF(s)\,G(t-s), \qquad F,G \in \Delta^+.$$

Here, the integral is the Lebesgue–Stieltjes convolution of F and G. The map τ satisfies the following properties:

- **Commutativity:** convolution is commutative.
- **Associativity:** convolution is associative.
- **Monotonicity:** τ is nondecreasing in each argument.
- **Identity:** ε_0 acts as a neutral element for convolution, since it is a point mass at 0.

Hence, τ is indeed a triangle function. The inequality (2) can then be written as $F_{pr}(t) \geq (\tau(F_{pq}, F_{qr}))(t)$, so the probabilistic triangle axiom holds for this model.

Example 3. Multi-locus genotype (hypercube). We consider a classical mutation model from population genetics.
State space. The state space consists of binary sequences of length L, $S = \{0,1\}^L$, where each sequence represents a genotype. The Hamming distance $d_H(x,y)$ counts the number of sites at which two genotypes differ.
Mutation model. Each site mutates independently according to a symmetric two-state continuous-time Markov chain: a flip $0 \leftrightarrow 1$ occurs at rate μ. Consequently, the global process is the product of L independent two-state CTMCs.
Exact-time transition probabilities. If the process starts at genotype x, the probability of being exactly at genotype y at time t is, by independence across sites,

$$P_x(X_t = y) = p_{\text{flip}}(t)^{d_H(x,y)}\, p_{\text{same}}(t)^{L-d_H(x,y)},$$

where for a single site with symmetric mutation rate μ,

$$p_{\text{same}}(t) = \frac{1+e^{-2\mu t}}{2}, \qquad \text{and} \quad p_{\text{flip}}(t) = \frac{1-e^{-2\mu t}}{2}.$$

This exact-time transition probability does not define a distance distribution in the sense of Δ^+, since it is not monotone in t (it typically increases initially and then decreases as the chain approaches stationarity). Nevertheless, it serves as the transition kernel from which hitting-by-time distributions can be constructed.

Lower bound via a chosen intermediate genotype z. For $L \geq 2$, a closed-form expression for $F_{xy}(t)$ is generally unavailable. Nevertheless, we can derive a useful lower bound and verify the triangle inequality as before. For any intermediate genotype z,

$$F_{xy}(t) \geq \int_0^t dF_{xz}(s)\, P_z(X_{t-s} = y).$$

The proof is identical to the single-locus case: condition on the first hitting time of z at time s, and then require the process to be at y during the remaining time interval of length $t-s$. This yields a lower bound, since the probability of being at y at some time after s is at least the probability of being at y exactly at time t.
Taking a supremum over z, or iterating the above bound by further integration, produces bounds that satisfy the convolution inequality. Consequently, choosing τ to be the (Lebesgue–Stieltjes) convolution again yields a valid triangle inequality.

Example 4. Take the single-locus case with $\mu = 0.1$(per unit time). Then $F_{Aa}(t) = 1 - e^{-0.1t}$. After $t = 10$ units, $F_{Aa}(10) = 1 - e^{1} \approx 0.6321$. So 63.2% chance to have mutated by time 10.
With three states in 2-state model we only have pairs, but we can combine chains in series and use convolution to compute bounds.
For a 4-locus genome ($L = 4$) with identical per-site rate $\mu = 0.1$, the probability to be exactly at a particular genotype y at time t if Hamming distance k from start is

$$P_x(X_t = y) = \left(\frac{1 - e^{-0.2t}}{2}\right)^k \left(\frac{1 + e^{-0.2t}}{2}\right)^{4-k}.$$

Here the single-site exponent is $2\mu = 0.2$ because two-state diagonalization gives $e^{-2\mu t}$.

Remark 2. *For CTMC mutation models, the hitting-by-time distribution $F_{ij}(t) = P_i(\exists\ s : X_s = j)$ belongs to Δ^+ (it is nondecreasing, left continuous, limits 0 and 1). (Shown explicitly for the 2-state case and argued generally using CTMC properties.)*

Using the strong Markov property and conditioning on first hitting times, the convolution inequality

$$F_{ik}(t) \geq \int_0^t dF_{ij}(s)\ F_jk(t-s)$$

holds for all states i, j, k. Therefore, choosing the triangle function τ to be the Stieltjes convolution, the probabilistic triangle axiom $F(i,k) \geq \tau(F(i,j), F(j,k))$ is satisfied.

2.3 Transition probability matrix of mutation processes

Let X_n denote the number of individuals of type A at time n. The state space may be finite or infinite. For each i, let $b_i > 0$ denote the probability that a single individual of type B mutates to type A, and let $d_i > 0$ denote the probability that a type A individual mutates to type B. Initially, we assume that at each time step exactly one birth and one death occur.

$$P_{ij} = p[X_{n+1} = j/X_n = i] = \begin{cases} b_i & j = i+1 \\ d_i & j = i-1 \\ 1 - b_i - d_i & j = i \\ 0 & j \neq i-1, i, i+1 \end{cases} \quad \text{for } i = 0, 1, 2, \ldots$$

For finite case $P_{N,N+1} = b_N = 0$ where N is the total size of the population.

The transition matrix is $\begin{bmatrix} b_0 & 1-b_0 & 0 & 0 & 0 & ... & 0 \\ d_1 & 1-b_1-d_1 & b_1 & 0 & 0 & ... & 0 \\ 0 & d_2 & 1-d_2-b_2 & b_2 & 0 & ... & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & \cdots & 1-d_N \end{bmatrix}$.

Lemma 1. *The above matrix is a valid transition probability matrix. i.e, $\sum_j P_{ij} = 1$ for all i.*

Proof. For $i = 0$, $\sum_j P_{ij} = 1 - b_0 + b_0 = 1$. For $i = 1$, $\sum_j P_{ij} = b_1 + 1 - (b_1 + d_1) + d_1 = 1$ and so on. For a fixed i, $\sum_j P_{ij} = b_i + 1 - (b_i + d_i) + d_i = 1$ for all i. □

3 Probabilistic metric space in population genetics

3.1 Mutation metrics on type A individuals

Consider a population consisting of N individuals. Each individual is of gene type A or gene type B. The state space is given by the number of type A individuals at a given time. A mutation of an individual from type A to type B is interpreted as a death, and the probability of a death in state i is denoted by d_i. Similarly, a mutation of an individual from type B to type A is interpreted as a birth, and the probability of a birth in state i is denoted by b_i.

Definition 5. *Let S be the state space $\{0,1,2,...\}$. P is the transition matrix for S that defines a mutation process $X_n, n \in N \cup \{0\}$. When this process is time homogeneous, for given two states i and j in S, define*

$$F_{ij}(t) = P[\text{ exist a time } s < t \text{ such that } X_s = j | X_0 = i]. \tag{3}$$

Remark 3. *Let $f_{ij}^{(s)}$ denote the probability that $X_0 = i$, the first time we visit state j is exactly at time s. Then $F_{ij}(t) = \sum_{s=1}^{u} f_{ij}^{(s)}, u < t$.*

Theorem 1. *Let S, P, $(X_n)_{n\in\mathbb{N}}$, and $F_{ij}(t)$ be defined as in the previous definition. Define the mapping $\mathscr{F} : S \times S \to \Delta$, where Δ denotes the space of cumulative distribution functions, by $\mathscr{F}(i,j) = F_{ij}$. Then the following properties hold:*

(i) $F_{ij}(0) = 0$;

(ii) F_{ij} *is left-continuous;*

(iii) $\mathscr{F}(i,i) = \varepsilon_0$ *for all* $i \in S$;

(iv) $\mathscr{F}(i,j) \neq \varepsilon_0$ *for* $i \neq j$;

(v) $\mathscr{F}(p,r) \geq \mathscr{F}(p,q)\mathscr{F}(q,r)$ *for all* $p,q,r \in S$,
where ε_0 *is given in* (1).

Proof. By (3), part (i) is obvious. For (ii), we have

$$F_{ij}(t-h) = P[\text{ exist a time } s < t-h \text{ such that } X_s = j | X_0 = i].$$

Then, $\lim_{h\to 0} F_{ij}(t-h) = P[\text{exist a time } \quad s < t : X_s = j | X_0 = i] = F_{ij}(t)$, i.e, F_{ij} is left continuous. For (iii), as $F_{ii}(t) = P[\text{exist a time } \quad s < t : X_s = i | X_0 = i]$ for $s = 0$, $X_s = i$ when $X_0 = i$. Therefore

$$F_{ii}(t) = \begin{cases} 1 & t > 0 \\ 0 & t \leq 0. \end{cases}$$

(iv) Let $i \neq j$, it needs a positive time s_0 to reach state j from the state i. For $0 \leq t \leq s_0 \quad F_{ij}(t) = 0$ and obviously $F_{ij}(t) \neq \varepsilon_0$.

(v) For the different states p,q and r and time points t_1 and t_2, we have

$$\begin{aligned}
F_{pr}(t_1+t_2) &= P[\text{there exist a time} \quad s<t_1+t_2 : X_s = r|X_0=p] \\
&\geq P[\text{there exist a time } i<t_1 : X_i = q \text{ for the first time and } X_s = r \text{ for some } i<s<t_1+t_2|X_0=p] \\
&= \sum_{i=1}^{[t_1]} P[X_i=q \;\text{ for the first time and }\; X_s=r \;\text{ for some }\; s\in[i,t_1+t_2)|X_0=p] \\
&= \sum_{i=1}^{[t_1]} P[X_i=q \text{ for the first time}|X_0=p]P[X_s=r \text{ for some } s\in[i,t_1+t_2)|X_0=p \text{ and } X_i=q \\
&\qquad\qquad \text{for the first time}] \\
&= \sum_{i=1}^{[t_1]} f_{pq}^i P[X_s=r \quad \text{for some} \quad s\in[i,t_1+t_2)|X_i=q \quad \text{for the first time}] \\
&= \sum_{i=1}^{[t_1]} f_{pq}^i P[X_s=r \quad \text{for} \quad s\in[0,t_1+t_2-i)|X_0=q] \\
&\geq \sum_{i=1}^{[t_1]} f_{pq}^i P[X_s=r \quad \text{for some} \quad s\in[0,t_2)|X_0=q] \\
&= F_{pq}(t_1)F_{qr}(t_2).
\end{aligned}$$

It means $\mathscr{F}(p,r)\geq\mathscr{F}(p,q)\mathscr{F}(q,r)$. □

Theorem 2. *Let $S,P,(X_n)_{n\in\mathbb{N}}$ and $F_{ij}(t)$ be defined as above. Then $(S,\mathscr{F},\tau)$ is a probabilistic metric space with a non-symmetric distance function where $(\tau(F,G))(x)=F(x)G(x)$, when $P(T_{ij}<\infty)=1$ where $T_{ij}=\{min(n\geq 0: X_n=j|X_0=i)\}$.*

Proof. It follows from Definition 4 and Theorem 1 □

3.2 Mutation metrics on type A subsets in finite populations

In the above model, each state represents the number of type A individuals in the population. However, different groups of type A individuals with the same cardinality are represented by the same number in this case. In the current construction, each distinct group of type A individuals is considered as a separate state. Initially, we assume that at most one birth and one death occur at a time. When the composition of the entire population is known, it is possible to determine the probability of a single birth, P_b, and a single death, P_d, at any given time point.

Definition 6. *Let P_b and P_d denote the probabilities of a single birth and a single death, respectively. At any time, the population of type A individuals can be represented as a subset of the whole population S, where $|S|$ denotes the total population size. For two subsets of type A individuals, S_i and S_j, we define the transition probability from S_i to S_j as $P_{ij}=P(X_{n+1}=S_j \mid X_n=S_i)$.*

Lemma 2. *Take $a_1 : S_j\subset S_i$ and $|S_i\backslash S_j|=1$, $a_2 : S_i\subset S_j$ and $|S_j\backslash S_i|=1$, $a_3 : |s_i|\neq 0$ or $|P|$, $a_4 : |S_0|=0$ or $|P|$ and $a_5 : S_i=S_j$. Then,*

$$P_{ij}=\begin{cases} P_d\frac{1}{|S_i|} & \textit{for} \quad a_1 \quad \textit{and} \quad a_3 \\ P_b\frac{1}{|P|-|S_i|} & \textit{for} \quad a_2 \quad \textit{and} \quad a_3 \\ \frac{1}{|P|} & \textit{for} \quad (a_1 \quad \textit{or} \quad a_2) \quad \textit{and} \quad a_4 \\ 1-(P_d+P_b) & \textit{for} \quad a_5 \quad \textit{and} \quad a_3 \\ 0 & \textit{otherwise} \end{cases}$$

leads to a transition probability matrix.

Proof. For a fixed i and $|S_i| \neq 0$ or $|P|$.

$$P_{ii} = 1-(P_d+P_b)$$

and

$$\sum_{j\neq i} P_{ij} = |S_i| P_d \frac{1}{|S_i|} + (|P|-|S_i|) P_b \frac{1}{(|P|-|S_i|)}$$
$$= P_d + P_b$$

(First term for $|S_i \backslash S_j| = 1$ and second term for $|S_j \backslash S_i| = 1$ and 0 for other cases)

$$\sum_j P_{ij} = 1-(P_d+P_b)+P_d+P_b = 1$$

Similarly we can prove for other cases. □

Definition 7. *Let $\mathscr{P}$ be the population and State space S is the power set of $\mathscr{P}$. Let P be the transition matrix for S which defines a time homogeneous Markov chain $(X_n)_{n\in\mathbb{N}}$. For any two given subsets S_i and S_j, we define*

$$F_{ij}^S(t) = P[\text{ there exist a time } s<t \text{ such that } X_s = S_j | X_0 = S_i]$$

Remark 4. *If f_{ij}^s denote the probability that with $X_0 = i$, the first time we visit j is exactly at time s, then*

$$F_{ij}^S(t) = \sum_{s=1}^{u} f_{ij}^s, u<t$$

Theorem 3. *Theorem 3 Let $S, P, (X_n)_{n\in\mathbb{N}}$ and $F_{ij}(t)$ be defined as in Definition 7, then $(S, \mathscr{F}, \tau)$ is a probability metric space with a non-symmetric distance function where $(\tau(F,G))(x) = F(x)G(x)$, when $P(T_{ij} < \infty) = 1$ where $T_{ij} = \{min(n \geq 0 : X_n = S_j | X_0 = S_i)\}$*

Proof. It follows from Theorem 2 and Definition 7, replacing i and j by S_i and S_j respectively. □

Theorem 4. *Let S_1 and S_2 be two states or subsets of $\mathscr{P}$. Let $t_a = |S_2 \backslash S_1|$ is the number of new births not in S_1 and $t_b = |S_1 \backslash S_2|$ be the number of deaths or the number of elements in S_1 but not in S_2. Let $t^* = t_a + t_b = |S_2 \backslash S_1| + |S_1 \backslash S_2|$. Then, $F_{12}(t) = 0$ for $t < t^*$. S_2 can be obtained from S_1 within a minimum of t^* steps.*

Proof. $t_a = |S_2 - S_1|$ and $t_b = |S_1 - S_2|$. Minimum number of steps or time to reach S_2 from $S_1 = t_a + t_b = t^*$ i.e, for $t < t^*$ it is impossible to reach S_2 from S_1 i.e, $F_{12}(t) = P[$ there exist a time $s < t^*$ such that $X_s = S_2 | X_0 = S_1] = 0$. □

3.3 Mutation metrics on a metric state space

In this case, we can consider the probabilistic metric space structure induced by the mutation process to probabilistically estimate the differences between states in the future due to changes in the initial state. Let $\mathscr{P}$ denote the population in a birth-death process, with the power set of $\mathscr{P}$ as the state space. A stationary Markov chain can then be constructed as in the previous section.

Definition 8. *Let $S_i, S_j \in S$ and $d(S_i, S_j) = |S_i \backslash S_j| + |S_j \backslash S_i|$. Then, (S, d) is a metric space. Now for each t, define $\mathscr{F}^{(t)} : S \times S \to \Delta^+$ as follows.*

$$\mathscr{F}^{(t)}(S_i, S_i)(x) = \varepsilon_0, \forall i$$

and

$\mathscr{F}^{(t)}(S_i, S_j)(x) = P\{$*distance between the states is less than x at time t, if the process starts at S_i and S_j*$\}$.

Using the results of Moynihan (1976), we obtain the following conclusions.

Theorem 5. *For each t, $(S, \mathscr{F}^{(t)}, \tau)$ is a pseudo probabilistic metric space under $\tau = \tau_{Tm}$.*

Proof. First, consider the case $S_i = S_j$. Then, $d(S_i, S_j) = 0$ implies that

$$\text{Prob}\{\text{distance between states} < x \text{ for } x \geq 0 \text{ at time } t \text{ when starting at } S_i \text{ and } S_j\} = 1 \text{ for } t \geq 0.$$

Hence, $\mathscr{F}^t(S_i, S_j) = \varepsilon_0$.
Next, consider the case $S_i \neq S_j$. Then $d(S_i, S_j) \neq 0$, but eventually the two states may reach the same state. Therefore, there exists a time t_0 such that the distance between the states can be zero, i.e.,

$$\mathscr{F}^{t_0}(S_i, S_j) = \varepsilon_0.$$

For fixed t, the left-continuity of $\mathscr{F}^t(S_i, S_j)$ follows from

$$\lim_{h \to 0} \mathscr{F}^t(S_i, S_j)(x - h) = \text{Prob}\{\text{distance between states} \leq x \text{ at time } t\} = \mathscr{F}^t(S_i, S_j)(x).$$

Symmetry is straightforward: for fixed t,

$$\mathscr{F}^t(S_i, S_j)(x) = \mathscr{F}^t(S_j, S_i)(x).$$

Finally, consider the triangle inequality for any $S_i, S_j, S_k \in S$ and $t \geq 0$:

$$\begin{aligned}
\mathscr{F}^t(S_i, S_k)(x_1 + x_2) &= \text{Prob}\{\text{distance between states at time } t < x_1 + x_2 \text{ starting at } S_i \text{ and } S_k\} \\
&\geq \text{Prob}\{\text{distance} < x_1 \text{ starting at } S_i, S_j \text{ and distance} < x_2 \text{ starting at } S_j, S_k\} \\
&\geq \mathscr{F}^t(S_i, S_j)(x_1) + \mathscr{F}^t(S_j, S_k)(x_2) - 1 \\
&\geq T_m\big(\mathscr{F}^t(S_i, S_j)(x_1), \mathscr{F}^t(S_j, S_k)(x_2)\big),
\end{aligned}$$

where T_m denotes the Lukasiewicz t-norm. Hence, for $\tau = T_m$,

$$\mathscr{F}^t(S_i, S_k) \geq \tau\big(\mathscr{F}^t(S_i, S_j), \mathscr{F}^t(S_j, S_k)\big).$$

This verifies all conditions for $(S, \mathscr{F}^{(t)}, \tau)$ to be a pseudo probabilistic metric space under $\tau = \tau_{Tm}$. $\square$

Theorem 6. *Let $\mathscr{F}_{ij}^{(t)} = \mathscr{F}^t(S_i, S_j)(x)$. Then for any real x, any states S_i, S_j, and any $r, t \geq 0$, we have*

$$F_{ij}^{(r+t)}(x) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} P_{im}(r) P_{jn}(r) F_{mn}^{(t)}(x),$$

or in matrix form,

$$F^{(r+t)}(x) = P(r) F^t(x) \big(P(r)\big)^T,$$

where $P(r)^T$ denotes the transpose of the matrix $P(r) = (P_{ij}(r))$.

Proof. By definition,

$$\mathscr{F}^t(S_i, S_j)(x) = F_{ij}^t(x) = \text{Prob}\{\text{distance between states} < x \text{ at time } t \text{ starting at } S_i \text{ and } S_j\}.$$

This can be expressed as

$$F_{ij}^t(x) = \sum_{k=0}^{\infty} P_{ik}(t) \sum_{\substack{l \\ d(l,k) < x}} P_{jl}(t).$$

For time $r+t$, we have

$$F_{ij}^{r+t}(x) = \sum_{k=0}^{\infty} P_{ik}(r+t) \sum_{\substack{l \\ d(l,k)<x}} P_{jl}(r+t). \tag{4}$$

Using the Chapman-Kolmogorov equation,

$$P_{ij}(r+t) = \sum_{k \in S} P_{ik}(r) P_{kj}(t),$$

we substitute into (4):

$$F_{ij}^{r+t}(x) = \sum_{k=0}^{\infty} \left[\sum_{m \in S} P_{im}(r) P_{mk}(t) \right] \sum_{\substack{l \\ d(l,k)<x}} \left[\sum_{n \in S} P_{jn}(r) P_{nl}(t) \right].$$

Interchanging the order of summation gives

$$\begin{aligned} F_{ij}^{r+t}(x) &= \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} P_{im}(r) P_{jn}(r) \left[\sum_{k=0}^{\infty} P_{mk}(t) \sum_{\substack{l \\ d(l,k)<x}} P_{nl}(t) \right] \\ &= \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} P_{im}(r) P_{jn}(r) F_{mn}^{t}(x), \end{aligned}$$

as claimed.

□

Acknowledgement

We express our gratitude to the reviewers for their thorough reading of the manuscript and their perceptive remarks.

References

Marcus, P. S. (1977). Probabilistic metric spaces constructed from stationary Markov chains. *Aequationes Mathematicae*, **15**, 169–171.

Moynihan, R. (1976). Probabilistic metric spaces induced by Markov chains. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **35**, 177–187.

Karlin, S. (1975). *A First Course in Stochastic Processes, Second edition.* Academic Press, New York.

Bhat, U. N., and Miller, G. K. (2002). *Elements of Applied Stochastic Processes* (Vol. 3). Wiley-Intersciencem, Hoboken, NJ.

Schweizer, B., and Sklar, A. (2005). *Probabilistic Metric Spaces*, Dover Publications, New York.

Murali, T. K., Vinod Kumar, P. B. and Pramod, P. K. (2022). Birth and death processes: a probabilistic metric space approach. Mathematics of Intelligent Computing and Data Science, Springer Proceedings in Mathematics& Statistics, Vol. 484, 223–235.

End of SMPS Journal – Volume 2, Issue 2

For manuscript submission and future issues, visit:
www.smps.um.ac.ir

Stochastic Models in Probability and Statistics

CONTENTS Vol. 2, No. 2, 2025

Web Site: https://smps.um.ac.ir
Email: smps@um.ac.ir
Print ISSN: 3115-9370
Online ISSN: 3060-7876