The role of consignees in empty container management

The role of consignees in empty container management

B. Legros, Y. Bouchery, J.C. Fransoo

Beta Working Paper series 517

BETA publicatie	WP 517 (working paper)
ISBN	
ISSN	
NUR	
Eindhoven	October 2016

The role of consignees in empty container management

Benjamin Legros[1] • Yann Bouchery[2] • Jan Fransoo[3]

[1] *Laboratoire Genie Industriel, CentraleSupélec, Université Paris-Saclay, Grande Voie des Vignes, 92290 Chatenay-Malabry, France*

[2] *Ecole de Management de Normandie, Laboratoire Métis, 30 Rue de Richelieu, 76087 Le Havre, France*

[3] *Eindhoven University of Technology, School of Industrial Engineering P.O. Box 513, 5600 MB Eindhoven, the Netherlands*

benjamin.legros@centraliens.net • ybouchery@em-normandie.fr • j.c.fransoo@tue.nl

Abstract

Nowadays, the majority of consumer goods is transported into a maritime container during at least one stage of the journey. Besides the many advantages of containerization, the management of empty containers is a key issue responsible for costly repositioning operations. This article investigates the potential for consignees to manage an inventory of empty containers at their location so as to enable direct reuse of these containers by shippers located in the surroundings. Our model aims at minimizing inventory holding and repositioning costs. These costs depend on the age of the container in the system in practice. We model the problem as a Markov decision process using the waiting time of the oldest container as a decision variable. Next, we use the value iteration technique to prove that a threshold policy in the age of the oldest container in stock is optimal. This paper, to the best of our knowledge, is the first to provide a proof for the optimality of a threshold policy based on the age of the oldest product in the system. We derive closed-form formulas for the system performance under the optimal policy. This allows us to compute numerically the optimal threshold and to derive explicit expressions of the optimal threshold in asymptotic regimes. We next analyze the impact of this proactive management of empty containers by the consignees on the level of direct container reuse. We show that this practice is very promising to enable a high level of direct reuse, but we also highlight that the consignees have little interest in implementing such a solution in the current setting of the container supply chain. We consequently propose remedies illustrated via a series of insights.

Keywords: Empty container management, Markov decision process, optimal inventory policy, value iteration technique, threshold.

1 Introduction

Containerization is the trend towards the deployment of a standard and dedicated international transport system for containers from door-to-door. This includes purpose-built containerships, dedicated deep-sea terminals with special handling equipment and dedicated intermodal infrastructure in the hinterland such as inland terminals. We refer to Levinson (2010) for a historical perspective on containerization. Containerization has shaped global supply chains by providing reliable, low cost and secure service for international trade. Container-based trade has expended from 50 million Twenty-foot Equivalent Units (TEUs) in 1996 to 180 million TEUs in 2015 (Trade and Development (2016)), such that containerized cargo now represents more than half the value of all international seaborne trade (Trade and Development (2016)). As a result, many authors consider ocean container transport as critical for global supply chain performance (Fransoo and Lee (2013)). As an example, Hausman et al. (2013) demonstrate the impact of logistics performance on trade by analyzing the time and costs of importing and exporting a container for 80 countries.

Container transport has many advantages including standardization, ease of handling, protection against damage and security. However, empty container movement is often considered as an important disadvantage of containerization. Indeed, once emptied at destination, the container often needs to be repositioned to be filled in again. Empty container repositioning occurs at the regional level within small geographical areas, but also at the global level between major seaports due to global trade imbalance (Boile et al. (2008)). For instance, the review of maritime transport published by the United Nations (2011) highlights that the costs of seaborne empty container repositioning was estimated to \$20 billion in 2009, while the costs of empty container repositioning in the hinterland was around \$10 billion for the same year. Globally, empty container repositioning accounted for 19% of the global industry income in 2009 (United Nations (2011)). We refer to e.g. Cheung and Chen (1998),Li et al. (2004), Olivo et al. (2005), Li et al. (2007), Lam et al. (2007), Erera et al. (2009), Long et al. (2015) for some studies on the global repositioning problem. For regional repositioning, Braekers et al. (2011) estimate that empty containers account for 40% to 50% of the regional movements.

In this article, we focus on empty container management in the hinterland of a deep-sea port, i.e., at the regional level. Besides the cost of moving empty boxes, empty container repositioning in the hinterland is responsible for many problems such as congestion in the port area and at deep-sea terminals, air pollution and accidents. As a consequence, a lot of research has been devoted to this issue. We refer to Dejax and Crainic (1987) for a review of early works in this field and we refer to Braekers et al. (2011) and Song and Dong (2015) for recent overviews. The main decisions considered deal with empty depots location (see e.g., Boile et al. (2008), Lei and Church (2011), Mittal et al. (2013)), container fleet sizing (see e.g., Du and Hall (1997)), inventory management for empty containers (see e.g., Yun et al. (2011)) and empty container routing (Jula et al. (2006), Deidda et al. (2008), Furió et al. (2013)).

Most hinterland regions have both import and export flows. These flows are not balanced in many cases and often very strongly imbalanced. Let consider a global import hinterland, i.e., a hinterland for which the import flows are greater than the export flows. Some containers will need to be back empty to the port, for global repositioning. However, some others will be reloaded with export cargo in the hinterland. For these containers, although inefficient, the common practice consists of having first the empty container being shipped back to the port, before being sent to another shipper where it is reloaded. This leads to situations such as the one described by Jula et al. (2006) for the ports of Los Angeles/Long Beach. In 2000, more than a million empty containers became available in the hinterland at local consignees' sites and were moved back empty to container terminals at one of the two ports. Meanwhile, 550,000 empty containers were transported from the two ports to local shippers. This situation appears in many major ports worldwide and leads to unnecessary movements of empty containers. Besides the financial impacts associated with this practice, unnecessary movements of empty containers lead to negative societal impacts such as pollution, congestion and accidents.

Braekers et al. (2011) highlight six strategies to reduce the unnecessary movements of empty containers in the hinterland, i.e., the use of inland container depots, street-turns, container substitution, container leasing, foldable containers and Internet-based systems. We refer to the Appendix 1 in Braekers et al. (2011) for a detailed description of these six strategies. This article focuses on the street-turn strategy as this one is often considered as one of the most efficient strategies for empty container management in the hinterland. The idea behind a street-turn is very basic and consists of shipping the empty container directly from the consignee to the shipper without passing by a terminal. The advantages of the street-turn strategy are numerous. The empty movements are reduced and accordingly repositioning costs decrease. Similarly, the reduction of empty movements may imply a decrease in the number road accidents. Each street-turn also avoids two movements to and from the terminal, reducing congestion. Finally, empty container demand from the shippers can be met sooner, increasing the container utilization rate which positively impacts container fleet sizing (Jula et al. (2006), Dong and Song (2012)).

Even if the advantages of the street-turn strategy appear to be important for all parties involved, Lei and Church (2011) highlight that street-turns are only used 10% of the time in the hinterland of Los Angeles/Long Beach. Identically, Wolff et al. (2007) found that the share of street-turns was in a range of 5-10% in the hinterland of the port of Hamburg. There are multiple explanations for the limited use of street-turns. First, operational limitations includes location and timing mismatch between the offer of empty containers at the consignees and the demand at the shippers, container type and ownership mismatch, lack of information sharing between consignees and shippers as well as the limited free time for container use by consignees and shippers (before incurring detention fees). Second, institutional barriers include the lack of procedure for such a direct interchange between a consignee and a shipper as well as inspection and paperwork issues. Finally, insurance issues also contribute to limit the generalization of street-turns (Jula et al. (2006)).

In this article, we aim at better understanding the operational limitations to street-turns. We especially focus on location and timing mismatch. This is in line with Song and Dong (2015) who consider that trade imbalance, dynamic operations and uncertainties are the critical factors explaining the high level of empty movements. Containers are generally controlled by the shipping lines who own and/or lease a pool of containers. As a consequence, the street-turn strategy is traditionally studied from a shipping line perspective as in Jula et al. (2006), Deidda et al. (2008) and Furió et al. (2013). As noticed by Braekers et al. (2011), the consignee is generally not recognized as a player involved in empty container management at the regional level, and therefore, empty containers that becomes available at a consignee's site are directly moved back to the port or to an empty depot. Indeed, managers often consider empty container movements as a necessary evil. This article analyzes the potential for consignees to become proactive in empty container management as we consider that they are at the best place to handle the container location and timing mismatch issues. Matching empty containers in the hinterland would indeed allow consignees to lower their empty container repositioning costs.

Ocean container transport in general has deserved a lot of attention from the transport and maritime economics communities. Many relevant and inspiring results have been generated. However, the transport and maritime economics literature does not take the inventory perspective into account, and hence their research paradigms cannot be used to tackle the problem we consider here. We model the problem as a Markov decision process and use the age of the oldest container as a decision variable. This Operations Management (OM) technique allows us to model the inventory of empty containers at a consignee's location. Examples of successful applications of OM in maritime transportation include container scheduling from a shipping line perspective (Choi et al. (2012)), ocean freight service purchasing by shippers (Lee et al. (2015)) and container inspections by customs (Bakshi and Gans (2010), Bakshi et al. (2011)). Several authors claim that identifying options for street-turns is a challenging task (see e.g., Braekers et al. (2011)). We firmly believe that OM techniques will contribute to a better understanding of the opportunities and impediments regarding street-turn strategies.

In this article, we investigate the potential for consignees to proactively manage empty containers at their location to enhance the feasibility of street-turns. From our knowledge, this article is the first one to investigate this option. Our contribution is twofold. First, we identify the optimal inventory policy as a threshold policy in the age of the oldest container. To the best of our knowledge, this article is the first to prove the optimality of a threshold structure based on the time spent in the system using the value iteration method. Next, we derive closed-form expressions of the performance measures. This allows us to compute numerically the optimal threshold. Moreover, using an asymptotic analysis we derive and evaluate closed-form expressions of the threshold. Second, we apply the model to several examples of inventory holding cost function of practical interest, and we derive a series of insights from the results obtained. We show that the proactive management of empty containers by consignees enable reaching a high level of street-turns in many cases. However, we show show that the consignees have little incentive to manage empty containers as the cost reduction obtained is often small due to detention fees. We investigate potential remedies such as an increase in the detention free period, the possibility for consignees to be exempted from detention fees for a small number of their containers, and a change in the detention fees structure and purpose.

We organize the rest of the article as follows. The setting we consider is described in Section 2. The policy analysis, including the proof of the optimal policy and the performance analysis is developed in Section 3. Section 4 is devoted to the analysis of special cases of inventory holding cost functions. This analysis enables us to study the behavior of the optimal inventory policy and the performance of this policy on the level of street-turns, in order to derive a series of practical insights. Finally, Section 5 is devoted to the conclusion and to future research directions. All proofs are given in the appendix at the end of the article.

2 Setting

Container transportation involves two main types of agreements between the shipping line and the consignee (respectively the consignor, depending on incoterms). Under carrier haulage, the shipping line provides door-to-door services, i.e., the shipping line takes responsibility of container transportation from port-to-port and in the hinterland. Under merchant haulage, the shipping line only takes care of port-to-port transportation. Inland transportation is controlled by the consignee (resp. consignor) who often contracts a logistics service provider. The shipping line charges detention and demurrage fees for the use of her container in the hinterland. Merchant haulage agreements are quite developed. For instance, Notteboom (2009) states that the average share of merchant haulage is 60% for the port of Rotterdam and 75% for the port of Antwerp. Similarly, Sterzik (2013) states that the average share of merchant haulage is about 70% in Europe. Accordingly, we focus on merchant haulage in what follows and we assume that the consignee is responsible for container transportation in the hinterland. This context is particularly relevant to study the role of consignees in empty container management.

After the laden container is released at the deep-sea port, the consignee arranges transportation and receives the container. Then the container is unloaded and needs to be sent back to the shipping line. In the vast majority of the cases in practice, the consignee decides to send the empty container back directly after being unloaded to the deep-sea terminal or to an empty depot. In this article, we investigate the option of keeping empties in inventory at the consignee's location until a shipper needing an empty container is identified, to achieve a street-turn. We exclude institutional barriers against street-turns as well as insurance issues from the analysis. We focus on a given type of container (e.g., 40 ft. dry containers) as container size is standardized. This leads to exclude container type mismatch from the analysis. Finally, due to concentration in the shipping line industry (Fransoo and Lee (2013)) and alliances (Agarwal and Ergun (2010)), we also exclude ownership mismatch from the analysis.

Under merchant haulage, the shipping lines indirectly control their containers in the hinterland by charging detention and demurrage fees. Demurrage fees are incurred when the container stays at the deep-sea terminal, and detention fees are charged when the container is in the hinterland, until being shipped back to the shipping line. Shipping lines want to get their containers back as quickly as possible, so they charge high detention and demurrage fees. As a consequence, consignees are sensitive to detention and demurrage fees and tend to negotiate detention and demurrage free periods. This leads to a complex structure in practice, with a free period, and several levels of fees. As an example, let us focus on a 40 ft. dry container imported to the port of Rotterdam by Maersk and assume that truck transportation is used to deliver the container to the consignee. The official detention fees according to Maersk (2016) include a free period of 3 days, then the rate is €55/day from day 4 to day 7 and finally the rate is €85/day after 7 days. In our discussions with consignees across Europe (Netherlands, France and Sweden), detention fees are always mentioned as one of the main barriers against street-turn strategies as the consignees feel that they do not have enough time to identify an export match before incurring high detention costs. Our model includes detention fees and we are consequently able to analyze the impact of detention fees on the level of street-turns.

We consider the regional empty container management problem from the consignee's perspective in isolation from the problem of managing the cargo inside the containers. Indeed, the inventory management decisions traditionally focus on the cargo and exclude the transportation packaging units. As a first attempt in analyzing the management of empty containers from the consignee perspective, we aim at focusing on models that do not affect the management of the cargo itself. The supply of empty containers is consequently not considered as a decision variable in our model and we consider this one as stochastic. We assume that the arrival process of empty containers is Poisson with parameter λ.

We assume that transportation costs in the hinterland are incurred by the consignee, i.e., the consignee bears the costs for shipping the laden container from the deep-sea terminal to the consignee's location (in accordance with the merchant haulage setting) as well as the costs for repositioning the empty container.

Two options are available. Either the empty container is shipped back to the shipping line (either to the deep-sea terminal or to an empty depot depending on the cheapest available option), and incurs a cost of c_{sl} monetary units per empty container, or the consignee manages to identify a shipper in need of an empty container in a nearby location, and incurs a cost of c_{st} monetary units per empty container. We assume that the transportation costs and time from the consignee to the shipper are lower than the transportation costs and time from the consignee to the shipping line (otherwise, the consignee has no interest in sending any empty container to the shipper). We consequently define the repositioning costs as $c_s = c_{sl} - c_{st}$ ($c_s > 0$). In case of a street-turn, the shipper will receive an empty container for free (as the consignee is paying for transportation) and quicker than a container obtained from the shipping line. We accordingly assume that the shipper will always accept the street-turn. Moreover, we assume that all the shippers of interest for the consignee are located in the same zone, such that transportation costs and time from the consignee to each feasible shippers are similar. We further assume that the aggregated demand for empty containers is stochastic and follows an exponential distribution with rate μ. Note that we refer to a single shipper for clarity reasons.

The objective for the consignee is to find the optimal empty containers inventory policy which minimizes the long-run expected costs. In case an empty container is kept in inventory, the consignee incurs linear inventory holding costs due to physical storage costs as well as detention costs if the detention free period is over. We aggregate these costs and we refer to them as inventory holding costs for simplicity. The inventory holding costs depend on the actual time spent by a container in the hinterland. The cost structure is not necessarily a linear function due to detention fees but it is often a more complex convex function. Although a state description based on the number of containers in the hinterland is common in Markov chain analysis, it does not allow to evaluate the overall cost per time unit of a set of containers. The information on the age of each container should be added in order to completely describe a given state of the system. This would make the state description very complex.

Since we assume that the inventory holding cost per time unit is an increasing and convex function of time, the oldest container in the hinterland is also the most costly container. Hence, if a decision to send back a container to the shipping line should be taken, this decision should be in priority taken for the oldest container. In other words, due to the increasing structure of the inventory holding costs, it is optimal to apply a first-in-first-out policy for sending back containers. For the same reason, a first-in-first-out discipline is optimal when a container can be sent to the shipper. Combining the Poisson arrival process with the first-in-first-out discipline, the age of the oldest container allows us to evaluate at any time the distribution of the containers as a function of their age. Therefore the age of the oldest container is chosen as a decision variable to determine the optimal policy.

We prove, in the setting considered, that the optimal control policy is a threshold policy on the age of the oldest empty container in inventory. We derive closed-form formulas for the performance measures. This allows us to obtain the optimal threshold. For several examples of inventory holding cost functions, we study the impact of the parameters on the performances of empty container management and we derive some practical insights.

3 Policy Analysis

In this section, we prove that the optimal control policy is a threshold policy on the age of the oldest container. Next, we compute the performance measures under this policy.

3.1 Optimal policy

We consider the set of all non-preemptive non-anticipating first-in-first-out policies for sending back containers to the shipping line. At any point of time, we want to decide for the oldest container (if any) whether to keep it, or to send it back to the shipping line. Whenever a match is possible it is optimal to realize the match. Since interarrival times of empty containers are exponential, the information on the oldest container allows us to estimate the distribution of the queue length (Koole et al. (2012)). Therefore the queue length does not need to be another decision variable because it completely depends on the age of the oldest container.

The objective function is composed by the inventory holding costs (including detention costs) and the costs of sending back a container to the shipping line. Concretely, the goal is to find the optimal policy which minimizes the following weighted sum which represents the long-run expected cost per container;

$$E(C) = c_s P_s + E(D), \tag{1}$$

where the coefficient c_s ($c_s \geq 0$) is the repositioning cost defined in Section 2 (Recall that c_s is the difference between the costs of sending back a container to the shipping line and the costs of shipping a container to the shipper), P_s is the long-run expected proportion of containers sent back to the shipping line and $E(D)$ is the expected inventory holding cost for a given container.

We propose to formulate the problem as a Markov decision process and next use the value iteration technique to prove the form of the optimal policy. We propose here a non-standard definition for the system states. We explicitly model the age of the oldest container, instead of the traditional modeling using the number of containers. This idea has been first proposed by Koole et al. (2012) in order to analyze queueing systems. The approach consists first of discretizing the age of the oldest container using successive exponential stages, each with rate γ, and next consists of reporting the waiting stage in the Markov process. The total number of stages required is not known beforehand. This is determined by the demand from the shipper. Having large values of γ improves the approximation as it better represents the continuously elapsing time. As γ tends to infinity, this approximate setup converges to the original one, which in turn leads to an exact analysis.

Let us denote by x a state of the system, where $x \geq 0$. State $x = 0$ corresponds to an empty inventory. States with $x > 0$ correspond to a situation where the oldest container in the system has an age of x stages. We denote the corresponding transition rate from state x to state x' by $t_{x,x'}$. Recall from Koole et al. (2012) that the transition probabilities denoted by $q_{x,x-h}$ from a stage x to a stage $x-h$, are $q_{x,x-h} = \left(\frac{\lambda}{\lambda+\gamma}\right)\left(\frac{\gamma}{\lambda+\gamma}\right)^h$

and $q_{x,0} = \left(\frac{\gamma}{\lambda+\gamma}\right)^x$ for $x > 0$ and $0 \leq h < x$. Hence, for $x, x' \geq 0$, we may write

$$t_{x,x'} = \begin{cases} \lambda, \text{ if } x = 0, x' = 1, \\ \gamma, \text{ if } x' = x+1, x > 0 \\ \mu q_{x,x-h}, \text{ if } x' = x-h, x > 0, \text{ and } 0 \leq h \leq x \\ 0, \text{ otherwise,} \end{cases}$$

which corresponds to arrival, demand and elapsing times.

Let us denote by $V_n(x)$ the expected cost over n steps, for $n \geq 0$ and $x \geq 0$. We pay a cost of c_s per container sent back to the shipping line and let $c(x)$ be the inventory holding cost function per container. We assume that $c(x)$ is a general increasing and convex function of the age the container. The cost $c(x)$ can therefore be used to model both the holding cost of storage and the detention cost which can include a detention free period. We choose to discretize our continuous-time model. This is possible because it is uniformizable (Section 11.5.2. in Puterman (1994)). The uniformization is done using the maximal event rate $\lambda + \mu + \gamma$, that we assume equal to 1. We denote by F the operator on the set of functions f from $\mathbb{N}$ to $\mathbb{R}$ defined by $F(f(x)) = \sum_{h=0}^{x} q_{x,x-h} f(x-h)$ for $x > 0$, and $F(f(0)) = f(0)$ for $x = 0$. For $n \geq 0$, we have

$$\begin{aligned} V_{n+1}(0) &= \lambda W_n(0) + (1-\lambda)V_n(0), \\ V_{n+1}(x) &= \gamma W_n(x) + \mu(F(V_n(x)) + c(x)) + (1-\gamma-\mu)V_n(x), \text{ for } x > 0, \end{aligned}$$

with

$W_n(x) = \max(F(V_n(x)) + c_s + c(x), V_n(x+1))$ if $x \geq 0$. We assume that $V_0 = W_0 = 0$.

For each $n > 0$ and every state x ($x \geq 0$) there is a minimizing action: send an empty container to the shipping line or keep all containers in the inventory. For fixed n ($n > 0$) we call this function:

$$\mathbb{N} \rightarrow \{\text{keep}, \text{send}\},$$

a policy. As n tends to infinity, this policy converges to the average optimal policy, that is, the policy that minimizes the long-run expected average costs (Puterman (1994)).

In Theorem 1, we prove that the optimal policy at iteration n for sending back containers is of threshold type under the conditions $\mu > \lambda$, $\gamma > \mu$ and $c(1) < c_s$. In Corollary 1, we show that in the long run, as γ tends to infinity, the optimal policy is also of threshold type without any restricting conditions on the system parameters. To the best of our knowledge, this paper is the first to prove the optimality of a threshold structure based on the time spent in a system using the value iteration technique. Most of continuous time Markov decision processes use the number in the system as a decision variable. This allows simple decisions but may not be optimal in a setting where the cost function substantially changes over time. Koole et al. (2012) provided a method to model the waiting time of a customer in a queue to allow for decisions based on the waiting but did not use the method to theoretically prove structural properties of a policy. Another numerical illustration of the method can be found in Koole et al. (2015).

Theorem 1 *For $\mu > \lambda$, $\gamma > \mu$, $c(1) < c_s$, the optimal policy for sending back containers at iteration n is of threshold type. There exists a threshold on the number of stages at which an empty container is sent back to the shipping line.*

Corollary 1 *As γ tends to infinity, the optimal long-run policy is a threshold policy on the age of the oldest empty container.*

The optimal routing result in Theorem 1 and Corollary 1 allow us to define more precisely the optimal rejection policy for empty containers. All containers are allowed to join the inventory, regardless of the system state. However, the system does not allow containers to infinitely stay in the inventory. A container waiting in at the consignee since exactly τ time units is sent back to the shipping line. Note that this policy is very simple to implement in practice.

3.2 Performance Analysis

We have proven in Section 3.1 that the optimal policy for the consignee is a threshold policy on the age of the empty containers. We evaluate here the performance of the system under this policy. We approximate the deterministic duration before rejection by an Erlang random variable with n stages and rate γ per stage. We choose n and γ such that $\frac{n}{\gamma} \triangleq \tau$. This ensures that as n and γ go to infinity, this Erlang random variable converges to the deterministic duration before rejection, τ.

We use the same state definition as in Section 3.1 except that the total number of stages in bounded. We denote by n the maximal number of stages. The transition structure is identical to the one in 3.1. The only difference is the transition from state $x = n$; a transition from n to a state $n - h$ can be caused not only by a μ-transition but also by a γ-transition which represents then a container sent back to the shipping line.

In Theorem 2, we give closed-form expressions for the probability of an empty system $p_0(\infty)$, the proportion of containers sent back to the shipping line, P_s, the expected time spent by a container in the inventory, $E(T)$, the expected number of container in the inventory, $E(N)$, and the proportion of empty containers which has spent more than t in the inventory, $P(T > t)$ with $0 \leq t \leq \tau$. These expressions are given as functions of the ratio $a = \lambda/\mu$. The ratio a represents the import/export balance.

Theorem 2 *We have:*

$$p_0(\infty) = \frac{1-a}{1-ae^{-\tau(\mu-\lambda)}}, \tag{2}$$

$$P_s = \frac{(1-a)e^{-\tau(\mu-\lambda)}}{1-ae^{-\tau(\mu-\lambda)}}, \tag{3}$$

$$E(T) = \frac{1-e^{-\tau(\mu-\lambda)}(1+a\tau(\mu-\lambda))}{\mu(1-a)(1-ae^{-\tau(\mu-\lambda)})}, \tag{4}$$

$$E(N) = a\frac{1-e^{-\tau(\mu-\lambda)}(1+a\tau(\mu-\lambda))}{(1-a)(1-ae^{-\tau(\mu-\lambda)})}, \tag{5}$$

$$P(T > t) = \mathbb{1}_{t<\tau}\frac{e^{-t(\mu-\lambda)} - ae^{-\tau(\mu-\lambda)}}{1-ae^{-\tau(\mu-\lambda)}}, \tag{6}$$

where $\mathbb{1}_{(x\in A)}$ *is the indicator function of a subset* A.

Our system reduces to an M/M/1+D queue where the arrival process is generated by the arrival of empty containers and the service is ensured by the shipper. The analysis to compute the performance measures is based on the discretization of the time spent by containers at the consignee's location. This method is an alternative to the first analysis of the M/M/1+D queue realized by Baccelli and Hebuterne (1981).

Cost function. In what follows, we derive the expression of the optimal cost function in the case where the inventory holding cost function is piecewise linear increasing (recall that this one includes detention fees). This enables us to obtain the optimal threshold τ and the related optimal performance measures.

We denote by c_k, the slope of the function on the intervals $[t_{k-1}, t_k)$ with the convention that $t_0 = 0$ for $1 \le k \le n$. The cost function is convex, therefore c_k is increasing in k. As a function of the time spent at the consignee, the cost function per container can be written as:

$$\begin{aligned} c(t) &= c_1 t\mathbb{1}_{t<t_1} + (c_2 t - t_1(c_2-c_1))\mathbb{1}_{t_1\le t<t_2} + (c_3 t - t_1(c_2-c_1) - t_2(c_3-c_2))\mathbb{1}_{t_2\le t<t_3} + \cdots \\ &= \sum_{n=1}^{\infty}\mathbb{1}_{t_{n-1}\le t<t_n}\left(c_n t - \sum_{k=1}^{n-1} t_k(c_{k+1}-c_k)\right). \end{aligned}$$

Using the results of Theorem 2, we obtain the expected cost per container;

$$E(C) = c_s + c(\tau)\frac{(1-a)e^{-\tau(\mu-\lambda)}}{1-ae^{-\tau(\mu-\lambda)}} + \int_{t=0}^{\tau}\mu\frac{(1-a)e^{-t(\mu-\lambda)}}{1-ae^{-\tau(\mu-\lambda)}}\cdot\sum_{n=1}^{\infty}\mathbb{1}_{t_{n-1}\le t<t_n}\left(c_n t - \sum_{k=1}^{n-1} t_k(c_{k+1}-c_k)\right)\mathrm{d}t.$$

We denote by n^* the integer such that $t_{n^*} \le \tau < t_{n^*+1}$, we thus have:

$$\begin{aligned} \int_{t=0}^{\tau}\sum_{n=1}^{\infty}\mathbb{1}_{t_{n-1}\le t<t_n}e^{-t(\mu-\lambda)}\left(c_n t - \sum_{k=1}^{n-1} t_k(c_{k+1}-c_k)\right)\mathrm{d}t &= \sum_{n=1}^{n^*}\int_{t=t_{n-1}}^{t_n} e^{-t(\mu-\lambda)}\left(c_n t - \sum_{k=1}^{n-1} t_k(c_{k+1}-c_k)\right)dt \\ &+ \int_{t=t_{n^*}}^{\tau} e^{-t(\mu-\lambda)}\left(c_{n^*+1} t - \sum_{k=1}^{n^*} t_k(c_{k+1}-c_k)\right)\mathrm{d}t \\ = \frac{-c(\tau)e^{-\tau(\mu-\lambda)}}{\mu-\lambda} + \frac{1}{(\mu-\lambda)^2}&\left(c_1 + \sum_{k=1}^{n^*} e^{-t_k(\mu-\lambda)}(c_{k+1}-c_k) - c_{n^*+1}e^{-\tau(\mu-\lambda)}\right). \end{aligned}$$

Finally, we obtain:

$$E(C) = c_s\frac{(1-a)e^{-\tau(\mu-\lambda)}}{1-ae^{-\tau(\mu-\lambda)}} + \frac{1}{\mu}\frac{c_1 + \sum_{k=1}^{n^*}(e^{-t_k(\mu-\lambda)} + at_k(\mu-\lambda)e^{-\tau(\mu-\lambda)})(c_{k+1}-c_k) - c_{n^*+1}e^{-\tau(\mu-\lambda)}(1+\tau a(\mu-\lambda))}{(1-a)(1-ae^{-\tau(\mu-\lambda)})}.$$

Structural results. In proposition 1, we give the first and second order monotonicty properties of the main performance measures as a function of the control parameter τ. These properties will be used to develop a method to compute numerically the optimal threshold for a general cost function. Moreover, these results will also be used in Section 4 to better explain the behavior of the expected cost.

Proposition 1 *For $a > 0$ and $\tau > 0$, the following holds:*

1. *The proportion of containers sent back to the shipping line, P_s, is strictly decreasing and strictly convex in τ (the proportion of matches is thus strictly increasing and strictly concave in τ).*

2. *The expected time spent at the consignee $E(T)$ and the expected number of containers are strictly increasing in τ.*

3. *The proportion $P(T > t)$ is strictly increasing in τ for $t < \tau$.*

Note that the expected time spent at the consignee as well as the expected number of containers at the consignee are neither convex nor concave in τ for $\tau > 0$. Note also that these monotonicity results apply for the the M/M/1+D queue. To the best of our knowledge these results can not be found in the queueing literature.

Optimal threshold. Using the monotonicity results, we develop a method to obtain the optimal threshold:

1. On each interval $[t_k, t_{k+1}]$ we compute $\frac{\partial E(C)}{\partial \tau}$;

$$\frac{\partial E(C)}{\partial \tau} = \frac{e^{-\tau(\mu-\lambda)}}{(1-ae^{-\tau(\mu-\lambda)})^2}\Bigg(-c_s\mu(1-a)^2 - c_1 a + c_{n^*+1}\left((1-a)(1+\tau a\mu) + e^{-\tau(\mu-\lambda)}(1+\tau a\mu(1-a)^2)\right) + \frac{1}{\mu}\sum_{i=1}^{k}(e^{-t_i(\mu-\lambda)} + t_i(\mu-\lambda)e^{-\tau(\mu-\lambda)}(1-2a))(c_{i+1}-c_i)\Bigg).$$

2. Either there exists a value for τ which is a zero of the above expression on the interval $[t_k, t_{k+1}]$ or we choose between t_k or t_{k+1} the value for τ which minimizes $E(C)$ on the interval. The value which would be a zero of $\frac{\partial E(C)}{\partial \tau}$ is unique.

3. We obtain a countable set of possible values for τ. Moreover, this set of values is finite due to the strict convexity of the cost function. Thus, as soon as increasing τ in a new interval only increases the expected cost the search for the optimal threshold finishes in the previous interval.

This method to obtain the optimal threshold can be related to other known problems in operations management like the optimal sizing of an inventory with discounted prices in Wilson's model. The optimal size in this model is either at the extremity of a segment or a zero of the derivative of the expected cost.

Asymptotic results. Due to the strong imbalance which can appear in an import zone ($a >> 1$) or in an export zone ($a << 1$) it might be interesting to derive asymptotic expressions of the performance measures. We write $f(a) \underset{a_0}{\sim} g(a)$ to express that $\lim_{a \to a_0} \frac{f(a)}{g(a)} = 1$, for $a_0 \in \mathbb{R}$.

In an import zone ($a >> 1$), we have:

$$P_s \underset{a\to\infty}{\sim} 1-\frac{1}{a}, \quad E(N) \underset{a\to\infty}{\sim} \lambda\tau, \quad P(T>t) \underset{a\to\infty}{\sim} \mathbb{1}_{t<\tau}\left(1-e^{(\tau-t)\mu(1-a)}\right).$$

In an export zone ($a << 1$), we have:

$$P_s \underset{a\to 0}{\sim} (1+\tau\lambda)e^{-\tau\mu}, \quad E(N) \underset{a\to 0}{\sim} a(1-e^{-\tau\mu}(1+\tau\lambda)), \quad P(T>t) \underset{a\to 0}{\sim} \mathbb{1}_{t<\tau}\left(e^{-t\mu}(1+t\lambda)\right).$$

These expressions will be used in the next section to obtain approximations of the optimal threshold.

4 When should we send back empty containers?

In this section, we consider different examples of inventory holding cost functions in order to gain insights on when to send back empty containers to the shipping line. We characterize situations for which the common practice that consists of sending back directly all containers (referred to as the immediate return policy) is optimal or not. We compare between the costs of our optimal policy to the costs obtained under the immediate return policy. We also analyze how our policy affects the proportion of street-turns.

4.1 Constant inventory holding costs per time unit

We consider here that the inventory holding cost per time unit is constant. This structure may apply if we exclude detention fees from the analysis. This will enable us assessing the impact of detention costs and complex detention costs structure on the performance of the system. The model may indeed also be used to assess the feasibility and effectiveness of implementing linear detention costs instead of the classical piecewise linear structure used. We denote by c_u the inventory holding cost of an empty container per unit of time. The key insight from Proposition 2 below is that the decision not to send back containers immediately to the shipping line only depends on the demand for empty containers; the flow of arriving containers does not influence the decision. More precisely, Proposition 2 proves that in case of a linear inventory holding cost function, our policy performs better than the immediate return policy if and only if $c_s\mu \leq c_u$. In other cases, the immediate return policy is optimal.

Proposition 2 *Two cases should be considered.*

1. *If $c_s\mu \leq c_u$, it is optimal to send back all empty containers to the shipping line (i.e., applying the immediate return policy),*
2. *If $c_s\mu > c_u$, there exists a unique finite threshold τ ($\tau > 0$) which minimizes the expected cost. This threshold is the unique solution of*

$$c_s\mu(1-a)^2 = c_u(1-2a+a(1-a)\mu\tau + a^2e^{-\mu(1-a)\tau}). \tag{7}$$

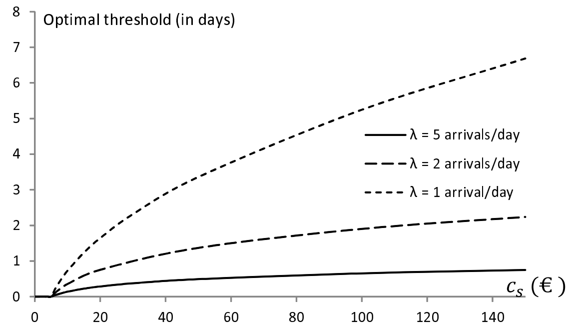
We now evaluate the impacts of c_s and c_u on the optimal policy. In proposition 3 and in corollary 2, we prove the impact of c_s and c_u on the optimal threshold τ and on the performance measures. Corollary 2 follows directly from combining proposition 1 and proposition 3.

Proposition 3 *The optimal threshold τ is increasing and concave in c_s and decreasing and convex in c_u.*

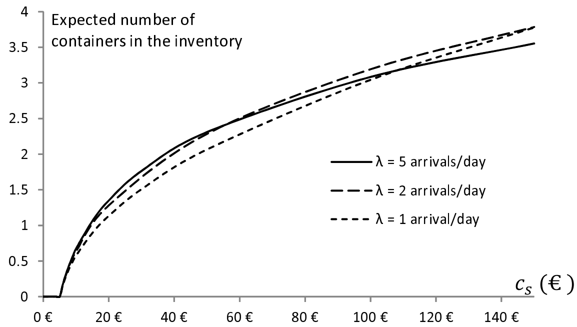
Corollary 2 *The proportion of matches, the expected number of empty containers at the consignee and the optimal cost per container are increasing and concave in c_s and decreasing and convex in c_u.*

Impact of c_s. In Figure 1, we illustrate the optimal policy as a function of c_s for a holding cost of €5 per day for different values of the arrival rate and an expected demand, μ, of 1 container per day. We take $a = 1$, $a = 2$ and $a = 5$ as examples. These values are representative of imbalance in major trade lanes. For instance, Theofanis and Boile (2009) report that trade imbalance for the transpacific lane increased from 50% (i.e., $a = 2$) to 67% (i.e., $a = 3$) from 2000 to 2005.

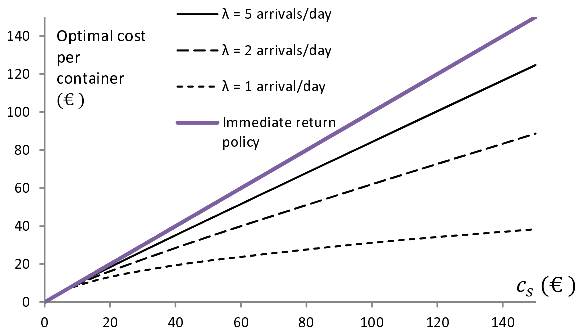
In Figure 1(a), we compute the optimal threshold, i.e., the maximal duration of a container at the consignee before being sent back. In Figure 1(b), we give the related optimal expected number of containers in the inventory. In Figure 1(c), we compute the optimal cost per container. As a comparison we also give the cost per container if the immediate return policy is implemented (note that in this case, the cost per container is equal to c_s). In Figure 1(d), we evaluate the proportion of matches. This proportion is computed as the ratio of the expected number of matches and the expected demand from the shipper on a given period of time. We indeed consider that $\lambda \geq \mu$ in our example so the maximum number of matches is limited by the demand for empty containers by the shipper.



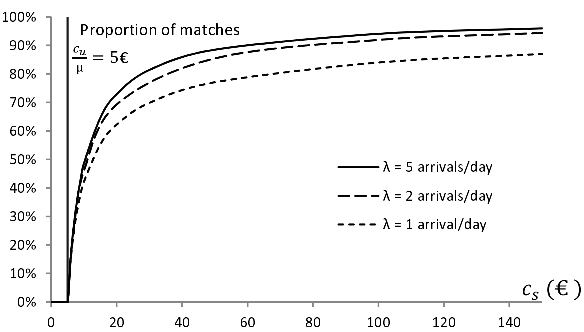
(a) Optimal threshold



(b) Expected number of containers in the inventory



(c) Optimal cost per container



(d) Proportion of matches

Figure 1: Optimal policy analysis ($\mu = 1$ container/day, c_u =€5/day)

We observe that:

1. As c_s increases (i.e., as the extra cost for sending back a container to the shipping line increases), τ

increases. This result is proven in Proposition 3 and Corollary 2. It is quite straightforward as an increase in c_s implies an increase in the expected benefit in case a match is realized.

2. As c_s increases, the average number of container in inventory increases (as a consequence of the increase in τ). This result directly follows from Proposition 1.

3. As c_s increases, the absolute and relative profit obtained when applying the optimal policy (as compared to the immediate return policy) increases. For reference, within the Netherlands, the cost c_s is about €100, which implies that the potential savings per container are from 16% for $\lambda = 5$ to 69% for $\lambda = 1$.

4. When omitting detention fees, the proportion of matches is very high as soon as $c_s\mu > c_u$. This means that empty container management by consignees shows a very high potential for reducing unnecessary movements of empty containers in the hinterland. Again, for the Dutch case with c_s =€100, note that the number of matches well exceeds 80% even for $\lambda = 1$. However, if $\lambda >> \mu$ the difference between the optimal cost per container and the cost in case of immediate return is very low. This means that the consignees do not have strong incentives to manage empty containers at their location.

5. In Figure 1(b), we can notice that the curves cross each other. So, if c_s is low the size of the inventory increases with λ and the opposite holds when c_s is high. This results from the competition of two phenomenons. First, for low arrival rates a relatively high number of containers is required in inventory in order to serve the relatively more volatile demand from the shipper. Second, for high arrival rates (with relatively lower demand volatility per unit of time) a high number of containers in inventory may be required in the inventory in order to reduce the flow of containers sent back to the shipping line with a cost c_s.

6. In case the demand for empty containers is held constant, the benefit of implementing the optimal policy decreases as the arrival rate of containers increases. In the opposite, the proportion of matches increases as the arrival rate of empty containers increases.

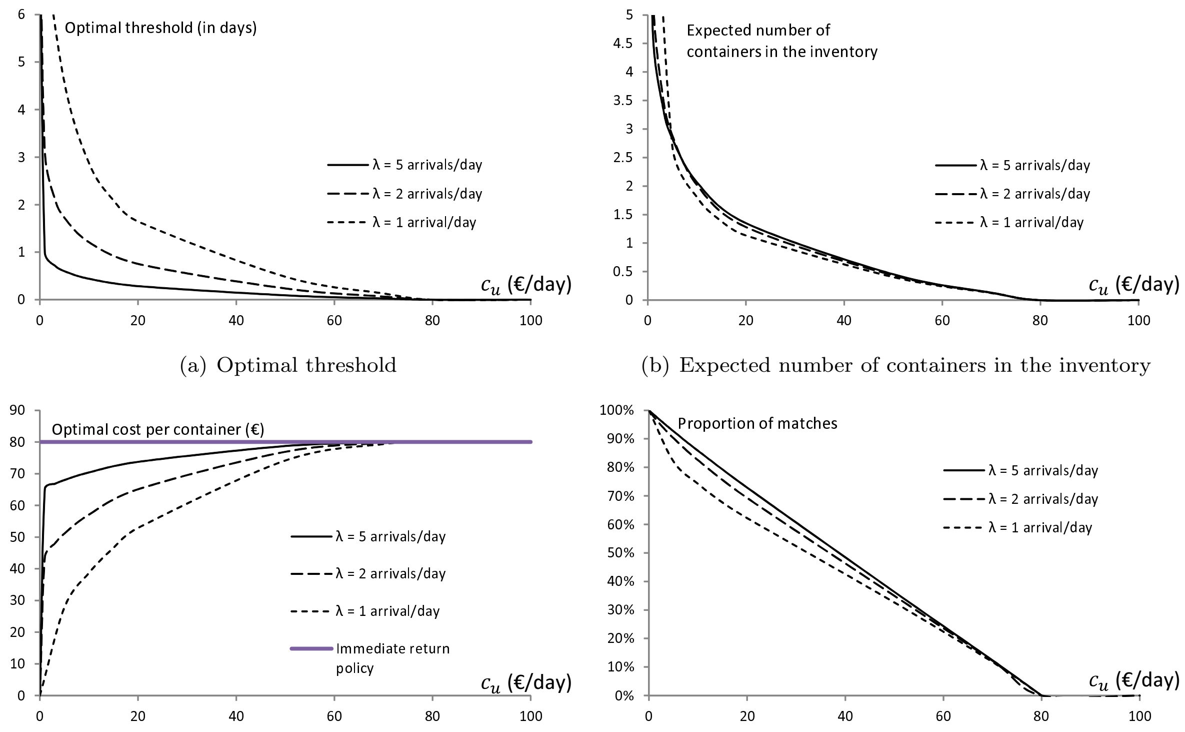
Impact of c_u. We illustrate the impact of c_u on the results with Figure 2. We set c_s =€80/container and μ =1 container/day. The results are computed as a function of c_u for different values of λ. Similarly to Figure 1, Figure 2(a) provides the optimal threshold in days, Figure 2(b) illustrates the resulting average number of containers in inventory. Figure 2(c) illustrates the optimal cost per container and Figure 2(d) focuses on the resulting number of matches.

We observe that:

1. As c_u increases, the optimal threshold as well as the average number of containers in inventory decreases (Figures 2(a) and 2(b)). This result is proven in Proposition 3 and Corollary 2.

2. The optimal cost per container is increasing in c_u. Additionally, due to the concavity in c_u (Corollary 2) the optimal cost per container is very sensitive to c_u for low values of c_u.

3. The proportion of matches is decreasing in c_u and the average number of matches is very sensitive to c_u. Additionally, for the parameter under study, the average number of matches is almost linearly



(a) Optimal threshold

(b) Expected number of containers in the inventory

(c) Optimal cost per container

(d) Proportion of matches

Figure 2: Optimal policy analysis ($\mu = 1$ container/day, $c_s =$€80/container)

deceasing in c_u, especially for high values of λ (Figure 2(d)). This reveals that the asymptotic behavior is almost reached. This observation differs from the convexity observed in Figure 2(a). One reason is that the proportion of matches is bounded whereas the number of containers is not. Another reason is that the expected number of containers involves both the first and the second order moments of the arrival and the demand process whereas the proportion of matches only involves the first moments. This explains also partly the stronger convexity of the number of containers.

4. As for Figure 1(b), we notice in Figure 2(b) that the curves cross each other. The reason is similar. If c_u is low then the size of the inventory decreases with λ and the opposite holds when c_u is high. In case of large values for λ, the probability of receiving a demand from the shipper before receiving a new empty container is quite small, so only a small number of containers in inventory ensures a high level of matches. Otherwise, more containers in inventory are necessary to obtain the same level of matches.

5. Again, as the arrival of containers increases, the benefits of implementing the optimal policy decreases and the proportion of matches increases.

Asymptotic analysis. Even if Equation (7) can be easily solved numerically, it does not lead to an explicit expression of the optimal threshold. We therefore additionally provide explicit expressions of the optimal threshold in extreme cases of the import/export balance. These expressions are derived from Equation (7) using Taylor expansions and equivalent expressions. The following holds in the case $c_s\mu > c_u$;

- As a tends to infinity (if the import/export balance is high), Equation (7) leads to

$$\tau \underset{a\to\infty}{\sim} \frac{1}{\lambda} \ln\left(\frac{c_s\mu}{c_u}\right).$$

- As a tends to zero (if the arrival of containers is low or if the demand is high), Equation (7) leads to

$$\tau \underset{a\to 0}{\sim} \frac{1}{\lambda} \left(\frac{c_s\mu}{c_u} - 1\right).$$

In Table 1, we evaluate the expected cost obtained using these two expressions of the threshold in comparison with the use of the optimal threshold for different values of the arrival rate. In the second and the third column we give the optimal threshold and its related expected cost. We compute the relative difference between the expected cost obtained with the two approximations and the optimal expected cost by rd= $\frac{E(C)_{\text{approximation}} - E(C)_{\text{optimal}}}{E(C)_{\text{optimal}}}$. As expected, Table 1 reveals that the first approximation is the best for high

Table 1: Performance comparison ($\mu = 1$, $c_s = 80$, $c_u = 5$)

λ	Optimal policy τ_{opt}	$E(C)_{opt}$	Approximation 1 $\tau_1 = \frac{1}{\lambda}\ln\left(\frac{c_s\mu}{c_u}\right)$	$E(C)_1$	rd$_1$	Approximation 2 $\tau_2 = \frac{1}{\lambda}\left(\frac{c_s\mu}{c_u} - 1\right)$	$E(C)_2$	rd$_2$
0.01	1485.00	0.05	277.26	0.05	0.00%	1500.00	0.05	0.00%
0.1	135.10	0.56	27.73	0.56	0.00%	150.00	0.56	0.00%
0.25	45.35	1.67	11.09	1.67	0.15%	60.00	1.67	0.00%
0.5	16.00	5.00	5.55	5.68	13.66%	30.00	5.00	0.02%
0.75	7.55	13.29	3.70	15.48	16.46%	20.00	14.70	10.54%
1	4.57	27.84	2.77	29.97	7.67%	15.00	44.84	61.08%
1.5	2.51	63.80	1.85	65.04	1.94%	10.00	105.54	65.42%
2	1.73	102.25	1.39	102.99	0.72%	7.50	150.04	46.74%
3	1.06	180.94	0.92	181.27	0.18%	5.00	232.50	28.50%
5	0.60	340.03	0.55	340.15	0.04%	3.00	393.75	15.80%
10	0.29	739.42	0.28	739.45	0.00%	1.50	794.44	7.44%
100	0.03	7938.92	0.03	7938.92	0.00%	0.15	7994.95	0.71%

arrival rate situations whereas the second one is the best for low arrival rate situations. However, it is interesting to observe that under low arrival rate situations the first approximation still performs well. The explanation is given by the result $\ln(1+x) \underset{x\to 0}{\sim} x$. Hence, the second expression found in the case where a tends to zero is equivalent to the first one found in the case where a tends to infinity. Note also that the worst degradation of the performance is observed with the second approximation.

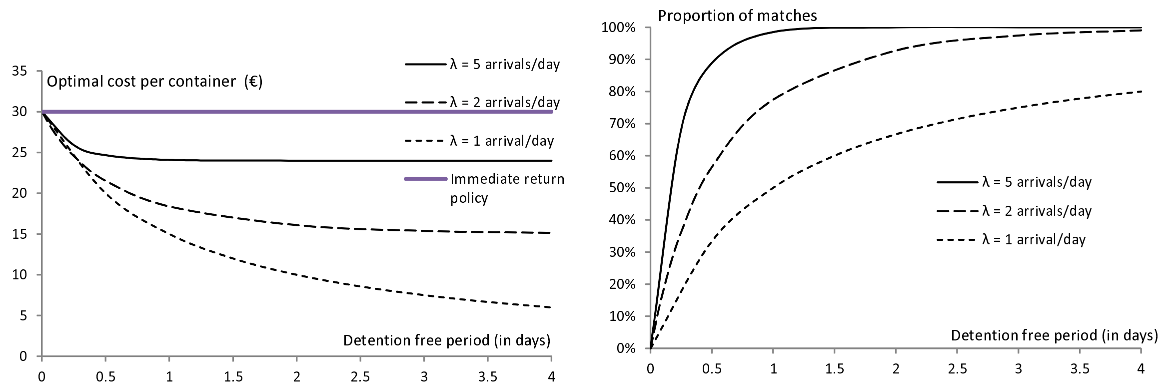
4.2 Impact of the detention free period

We assume here that the cost function is a simple one-step function; during a first detention free period no cost are encountered per container at the consignee. At the end of the detention free period, the cost is assumed to be infinite. This choice does not correspond to a realistic cost modeling but this corresponds to the behavior of a consignee that refuses to pay any detention fees, while neglecting the physical cost of keeping empty containers (the latter is often very low as the consignees often have some space available for staking few empties). This behavior is very common in practice according to the discussions we had with consignees in several European countries. In the setting we consider in this section, the optimal threshold is exactly equal to the duration of the detention free period.

The expected cost per container is therefore:

$$E(C) = c_s \frac{(1-a)e^{-\tau(\mu-\lambda)}}{1-ae^{-\tau(\mu-\lambda)}}.$$

In Figure 3, we present the impact of the duration of the detention free period on the difference in cost per container between the optimal policy and the immediate return policy (Figure 3(a)) and on the proportion of matches (Figure 3(b)). Note that we take $c_s =$ €30/container and $\mu = 1$ container/day through this example.



(a) Optimal cost per container

(b) Proportion of matches

Figure 3: Optimal policy analysis ($\mu = 1$ container/day, $c_s =$ €30/container)

We observe that:

1. As expected, the optimal cost per container is decreasing and convex in the detention free period, while the proportion of matches is increasing and concave in the detention free period. This observation is proven in Corollary 3.

2. Empty container management by the consignees has a very high potential for reducing the unnecessary movements of empty containers in the hinterland as soon as the residual detention free period at the consignee is high enough.

3. For high values of λ, the difference between the optimal cost per container and the cost in case of immediate return is very low. This means that the consignees do not have strong incentives to manage empty containers at their location in such a setting. This observation confirms the ones made in Section 4.1 with constant inventory holding costs.

4. In the case of a small λ, the proportion of matches is strongly affected by the detention free period for small values of τ. This result directly follows from Proposition 1.

Corollary 3 gives the impact of the detention free period on the optimal cost per container and the proportion of matches. The proof of corollary 3 directly follows from proposition 1.

Corollary 3 *The following holds:*

- *The optimal cost per container is decreasing and convex in the detention free period.*
- *The proportion of matches is increasing and concave in the detention free period.*

4.3 Managerial insights.

The theoretical results and numerical tests presented in this article enable us to derive a series of insights on how to enhance empty container reuse in the hinterland by focusing on street-turn strategies. We propose a new model that enables consignees to optimally handle a small amount of empties in order to identify possible export matches. The insights derived below aim at assessing the impacts of such proactive management of empty containers by consignees.

Insight 1 *The proactive management of empty containers by the consignees enables reaching a high level of street-turns in many cases. Street-turns are beneficial not only for the consignees (as a way to decrease their costs) but also for the shipping lines (by increasing the utilization rate of the containers and reducing the cost of sending back empty containers to shippers). Street-turns additionally enable reducing congestion, accidents and pollution.*

Insight 1 is directly derived from the observations made about the level of matches in Sections 4.1 and 4.2. Insight 1 is very important as this shows that the new management practice studied in this article may strongly help in optimizing container flows in the hinterland. However, Insight 2 shows that the consignees may not be willing to investigate the idea of managing empties at their location unless they are incentivized to do so in the current setting of many hinterlands.

Insight 2 *In case of a strong imbalance in favor of import between import and export flows and short detention free periods, the consignees have little incentive to manage empty containers at their location as the difference in cost as compared to the immediate return policy is small.*

Insight 2 is derived from the observations made in Sections 4.1 and 4.2 about the optimal costs per container as compared to the costs in case the immediate return policy is implemented. This second insight may be quite disappointing at first glance. Indeed, combined with Insight 1, we face a very effective management practice to handle the problem of reducing empty movements of containers in the hinterland, but we show that this practice is not very likely to be implemented. However, our results help in identifying how shipping lines may influence the behavior of the consignees and incentivize them to manage a small number of empty containers at their location. We consequently take the shipping line perspective in what follows.

First, we assess if detention fees are efficient in controlling the time spent by containers in the hinterland. We highlight in Section 2 that detention fees are often used as a way to ensure that containers are quickly shipped back to the shipping line. Now assume that we omit detention fees. Many shipping line companies would be reluctant to implement such type of tariff as they would be afraid of losing control of their containers, by incentivizing consignees to keep empties for a long time. The results we obtained in Section 4.1 hold. Figure 1(a) shows that optimal threshold is quite low, and even very low for consignees with high demand. Moreover, Figure 2(a) shows that this optimal threshold is very sensitive to c_u. This means that the consignees have little or no interest in keeping empty containers at their location as soon as they incur even little inventoy holding costs. We can conclude that empty containers will be shipped back to the shippers quite quickly, independently of detention fees as the physical costs of keeping empties, even if they are small, are sufficient to ensure a low value for the optimal threshold. We need to acknowledge here that we assume that consignees unload their containers directly after arrival, i.e., they do not use containers as cheap storage locations. From our knowledge, this practice is quite uncommon in Europe and North America. However, our study does not enable us to assess if detention fees really deter consignees to do so. We exclude this practice from the analysis in what follows and we derive the following insight.

Insight 3 *Detention fees are not necessary to ensure that containers do not spend too much time in the hinterland. The physical cost of storing empties is enough to deter consignees to keep empties for a long time before shipping them back to the shipping line.*

Insight 3 questions the real objective of the detention fees. Indeed, detention fees are often claimed to help shipping lines to control their containers in the hinterland, but they may also be a source of additional revenue for shipping lines. In what follows, we consequently investigate two scenarios. At first, we assume that shipping lines mainly focus on container utilization, and therefore, we propose a solution that may help in increase container reuse, at the risk of a decrease in detention fees collected from the consignees. In the second case, we will show how the detention fee structure can be modified to encourage consignees to manage empty containers at their location, while protecting revenues generated from detention fees.

First, we address the first scenario, for which the revenue generated by detention fees does not need to be hold constant. In this case, the shipping lines could propose to increase the detention free period for consignees who accept to hold few empties at their location to proceed to street-turns. This solution will not strongly affect the time spent by containers in the hinterland (as we show in Insight 3 that detention fees are not necessary to control the time spent by containers in the hinterland) and may be perceived by consignees as a strong incentive to proactively manage empties at their location. Also note that this solution may help shipping lines to reduce their costs for sending empty containers to the shippers. Especially, Figure 3(b) shows that in case the trade imbalance is not very high, an increase in the detention free period will have a strong effect on the level on street-turns.

Insight 4 *In order to increase the proportion of street turns in the hinterland, shipping lines may propose an increase in the detention free period for consignees who accept to hold few empties at their location.*

Insight 4 may help the shipping line to identify commercial solutions to entice consignees to proactively manage empty containers. This practice is already in place to entice consignees to use intermodal transportation in the hinterland. As an example, we stated in Section 2 that the detention free period for a 40ft. dry container imported to the port of Rotterdam and transported by truck was 3 days. Maersk (2016) additionally states that this detention free period is extended to 5 days in case of barge or train transportation. The same type of agreement may therefore be put in place to entice consignees to manage empties at their location. However, the optimal threshold level depends on λ. This means that the optimal detention free period would be specific to each consignee. To overcome this issue, we may notice that the expected number of container in inventory is quite similar for different values of λ according to Figure 1(b). This suggests an alternative strategy that allows for a given maximum number of empties to be stored free of detention fees at the consignees' location. As stated above, c_s =€100 in the Dutch case. According to Figure 1(b), this would mean that the shipping line should allow Dutch consignees to keep up to three empty containers at their location. Interestingly, this number is more or less independent of the average demand. This result is formulated in the following insight.

Insight 5 *In order to increase the proportion of street turns in the hinterland, shipping lines may allow consignees to keep up to a given number of empties at their location without incurring any detention fees. This will have a strong impact on the level of matches without compromising the control of empty containers by the shipping line.*

We investigate further the option of proposing a linear detention fee structure, by changing the aim of these fees. Indeed, detention fees are nowadays perceived as penalty in case of late delivery of empty containers to the shipping line. We could have them to be perceived as renting fees, for the equipment shipping lines provide to the consignee in case of merchant haulage. Assume, for instance, that the shipping line proposes a new detention tariff which consists of a single rate of €15/day (independently of the time spent by the container in the hinterland). Assume that this rate have been estimated to generate the same revenue as under the complex detention fee structure exposed in Section 2. The model of Section 4.1 can be used in this case as well, if we take c_u =€20/day (€15 detention fee + €5 inventory holding costs). Figure 2(d) shows that the proportion of match is quite high in this case, so this linear detention fee structure can be helpful in improving the direct reuse of containers in the hinterland. Of course, this may not work if the rate chosen is too high, but we can expect that the linear rates proposed by shipping lines would be quite low if they aim at equalizing with the revenues currently generated. To better investigate this statement, we illustrate a decomposition of the costs for Figure 2(b) in Figure 4 . We can notice that the detention cost per container is concave in the linear detention rate. Indeed, τ is decreasing in c_u as stated in Proposition 3, so when increasing the linear detention rate, more revenue is generated per container per time unit but the containers stay at the consignee for a shorter period of time. This analysis enables us to highlight that there exists a rate that maximize the profit from detention fees for the shipping line. We can notice that this rate is equal to €25/day for Figure 4, if we still consider a physical inventory holding cost of €5/day. For this detention rate, the proportion of matches is higher than 50% according to Figure 2(d). This enables us to derive the following insight.

Insight 6 *Proposing linear detention fees would lead to an increase in the proportion of street-turns as compared to short detention free periods if the consignees accept to manage empty containers at their location.*

Linear detention fees would at first be difficult to implement as consignees are used to detention free period, but we expect that relabeling them into renting fees would help. Overall, our results show that the detention fee structure used by shipping lines may be overly complex as compared to the results targeted (if we consider primarily the control of containers). In addition, this complex structure is sometimes criticized

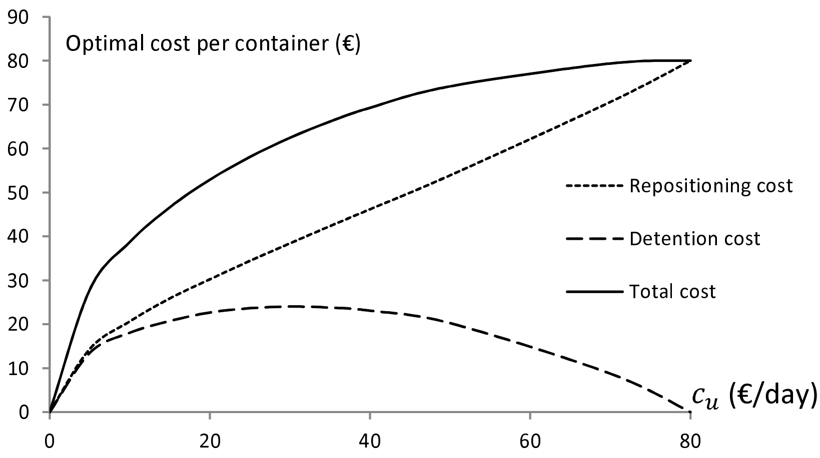


Figure 4: Decomposition of the costs in Figure 2(c) for $\lambda = 1$ arrival/day

as this one may induce unexpected behavior (see e.g. Fazi (2014)).

5 Conclusion

Street-turns are considered as one of the most efficient strategies for empty container management in the hinterland. In this article, we investigate if the proactive management of empty containers by consignees could lead to strongly improve the proportion of street-turns in the hinterland. From a level of street-turns of around 10% reported in the literature, our results show that much higher levels could be achieved if the consignees were proactive in managing empty containers. We propose a model of empty container management at the consignee's location. We formulate the problem as a Markov decision process using the age of the oldest container as a decision variable. Next, we identify and prove the structure of the optimal policy using a value iteration technique. This one is of threshold type in the age of the oldest empty container in the system. This means that all containers are allowed to join the inventory, regardless of the system state. However, the containers will be sent back to the shipping line if they are not used for a match after a given time period. This policy is original compared to many common practice policies since it is based on the age of the containers instead of the number of containers in the inventory. Moreover, it is simple to implement for managers. This paper, to the best of our knowledge, is the first to provide a proof for the optimality of a threshold policy based on the age of the oldest product in the system.

We next derive closed-form formulas for the performance measures under the optimal policy. This allows us to derive numerically the optimal threshold and to evaluate explicit expressions of the threshold in asymptotic regimes. We show that under realistic parameters value, the optimal proportion of street-turns is very high. This shows that the proactive management of empty containers by consignees may be very impactful. However, the difference between the total costs per container incurred by the consignee as compared to the costs incurred under the immediate return policy is low. This means that the consignees do not have strong incentive to manage empty containers at their location. We accordingly study potential remedies by investigating how shipping lines may modify their detention fees and/or detention fee structure to incentivize the consignees. These are preliminary ideas to solve this issue but many other ones may be investigated, such as the sharing of transportation costs between the consignee and the shipper in case of a street-turn. Additionally, other actors in the hinterland such as terminal operators, port authorities and local policy makers may entice consignees to investigate this option. We hope that this article will draw attention on empty container management by the consignees. This may indeed be one of the most powerful and simple option for tackling the problem of empty container repositioning in the hinterland.

References

Agarwal, R. and Ergun, Ö. (2010). Network design and allocation mechanisms for carrier alliances in liner shipping. *Operations Research*, 58(6):1726–1742.

Baccelli, F. and Hebuterne, G. (1981). On queues with impatient customers. *Performance'81*. North-Holland Publishing Company, 159-179.

Bakshi, N., Flynn, S. E., and Gans, N. (2011). Estimating the operational impact of container inspections at international ports. *Management Science*, 57(1):1–20.

Bakshi, N. and Gans, N. (2010). Securing the containerized supply chain: analysis of government incentives for private investment. *Management Science*, 56(2):219–233.

Boile, M., Theofanis, S., Baveja, A., and Mittal, N. (2008). Regional repositioning of empty containers: Case for inland depots. *Transportation Research Record: Journal of the Transportation Research Board*, (2066):31–40.

Braekers, K., Janssens, G. K., and Caris, A. (2011). Challenges in managing empty container movements at multiple planning levels. *Transport Reviews*, 31(6):681–708.

Cheung, R. K. and Chen, C.-Y. (1998). A two-stage stochastic network model and solution methods for the dynamic empty container allocation problem. *Transportation science*, 32(2):142–162.

Choi, B.-C., Lee, K., Leung, J. Y.-T., Pinedo, M. L., and Briskorn, D. (2012). Container scheduling: Complexity and algorithms. *Production and Operations Management*, 21(1):115–128.

Deidda, L., Di Francesco, M., Olivo, A., and Zuddas, P. (2008). Implementing the street-turn strategy by an optimization model. *Maritime Policy & Management*, 35(5):503–516.

Dejax, P. J. and Crainic, T. G. (1987). Survey paper-a review of empty flows and fleet management models in freight transportation. *Transportation Science*, 21(4):227–248.

Dong, J.-X. and Song, D.-P. (2012). Quantifying the impact of inland transport times on container fleet sizing in liner shipping services with uncertainties. *OR spectrum*, 34(1):155–180.

Du, Y. and Hall, R. (1997). Fleet sizing and empty equipment redistribution for center-terminal transportation networks. *Management Science*, 43(2):145–157.

Erera, A. L., Morales, J. C., and Savelsbergh, M. (2009). Robust optimization for empty repositioning problems. *Operations Research*, 57(2):468–483.

Fazi, S. S. (2014). *Mode selection, routing and scheduling for inland container transport*. PhD thesis, Technische Universiteit Eindhoven.

Fransoo, J. C. and Lee, C.-Y. (2013). The critical role of ocean container transport in global supply chain performance. *Production and Operations Management*, 22(2):253–268.

Furió, S., Andrés, C., Adenso-Díaz, B., and Lozano, S. (2013). Optimization of empty container movements using street-turn: Application to valencia hinterland. *Computers & Industrial Engineering*, 66(4):909–917.

Hausman, W. H., Lee, H. L., and Subramanian, U. (2013). The impact of logistics performance on trade. *Production and Operations Management*, 22(2):236–252.

Jula, H., Chassiakos, A., and Ioannou, P. (2006). Port dynamic empty container reuse. *Transportation Research Part E: Logistics and Transportation Review*, 42(1):43–60.

Koole, G., Nielsen, B., and Nielsen, T. (2015). Optimization of overflow policies in call centers. *Probability in the Engineering and Informational Sciences*, 29(03):461–471.

Koole, G., Nielson, B., and Nielson, T. (2012). First in line waiting times as a tool for analysing queueing systems. *Operations Research*, 60(5):1258–1266.

Lam, S.-W., Lee, L.-H., and Tang, L.-C. (2007). An approximate dynamic programming approach for the empty container allocation problem. *Transportation Research Part C: Emerging Technologies*, 15(4):265–277.

Lee, C.-Y., Tang, C. S., Yin, R., and An, J. (2015). Fractional price matching policies arising from the ocean freight service industry. *Production and Operations Management*, 24(7):1118–1134.

Lei, T. L. and Church, R. L. (2011). Locating short-term empty-container storage facilities to support port operations: A user optimal approach. *Transportation Research Part E: Logistics and Transportation Review*, 47(5):738–754.

Levinson, M. (2010). *The box: how the shipping container made the world smaller and the world economy bigger*. Princeton University Press.

Li, J.-A., Leung, S. C., Wu, Y., and Liu, K. (2007). Allocation of empty containers between multi-ports. *European Journal of Operational Research*, 182(1):400–412.

Li, J.-A., Liu, K., Leung, S. C., and Lai, K. K. (2004). Empty container management in a port with long-run average criterion. *Mathematical and Computer Modelling*, 40(1):85–100.

Long, Y., Chew, E. P., and Lee, L. H. (2015). Sample average approximation under non-iid sampling for stochastic empty container repositioning problem. *OR Spectrum*, 37(2):389–405.

Maersk (2016). Demurrage and detention tariff, port of rotterdam.

Mittal, N., Boile, M., Baveja, A., and Theofanis, S. (2013). Determining optimal inland-empty-container depot locations under stochastic demand. *Research in Transportation Economics*, 42(1):50–60.

Notteboom, T. E. (2009). Complementarity and substitutability among adjacent gateway ports. *Environment and Planning A*, 41(3):743–762.

Olivo, A., Zuddas, P., Di Francesco, M., and Manca, A. (2005). An operational model for empty container management. *Maritime Economics & Logistics*, 7(3):199–222.

Puterman, M. (1994). *Markov Decision Processes*. John Wiley and Sons.

Song, D.-P. and Dong, J.-X. (2015). Empty container repositioning. In *Handbook of Ocean Container Transport Logistics*, pages 163–208. Springer.

Sterzik, S. (2013). *Concepts, Mechanisms, and Algorithms to Measure the Potential of Container Sharing in Seaport Hinterland Transportation*. PhD thesis, Staats-und Universitätsbibliothek Bremen.

Theofanis, S. and Boile, M. (2009). Empty marine container logistics: facts, issues and management strategies. *GeoJournal*, 74(1):51–65.

Trade, U. N. C. and Development (2016). Review of maritime transport.

United Nations (2011). *Review of Maritime transport*. United Nations Publication, Geneva.

Wolff, J., Herz, N., and Flamig, H. (2007). Report on case study: Empty container logistics: Hamburg-baltic sea region. *Hamburg University of Technology, The Baltic Sea region programme*, 2013.

Yun, W. Y., Lee, Y. M., and Choi, Y. S. (2011). Optimal inventory control of empty containers in inland transportation system. *International Journal of Production Economics*, 133(1):451–457.

Appendix

Proof of Theorem 1

We prove by induction that the optimal policy for sending back containers is of threshold type. We thus need to show that $V_n(x+1) - F(V_n(x)) - c_s - c(x)$ is increasing in x, for $x \geq 0$ and $k \geq 0$. In other words, we need to show that

$$F(V_n(x)) + V_n(x+2) + c(x) \geq F(V_n(x+1)) + V_n(x+1) + c(x+1), \tag{8}$$

for $x \geq 0$ and $n \geq 0$. This relation is refereed to as *generalized convexity*. In this proof, we also have to show that V_n is increasing in x, for $x \geq 0$ and $n \geq 0$. Since $V_0(x) = 0$, then V_0 is increasing and generally convex (*igcv*).

First, we assume that V_n is *igcv* for a given $n \geq 0$, and we want to show that the same property holds for W_n. We show that if V_n is increasing in x, then $F(V_n)$ is also increasing in x. To simplify the notations, we denote by u the ratio $\frac{\lambda}{\lambda+\gamma}$. We have for $x \geq 0$,

$$\begin{aligned}
F(V_n(x+1)) - F(V_n(x)) &= \sum_{h=0}^{x+1} q_{x+1,x+1-h} V_n(x+1-h) - \sum_{h=0}^{x} q_{x,x-h} V_n(x-h) \\
&= \sum_{h=0}^{x-1} u(1-u)^h (V_n(x+1-h) - V_n(x-h)) - (1-u)^x V_n(0) + (1-u)^{x+1} V_n(0) + u(1-u)^x V_n(1) \\
&= \sum_{h=0}^{x-1} u(1-u)^h (V_n(x+1-h) - V_n(x-h)) + u(1-u)^x (V_n(1) - V_n(0)) \geq 0.
\end{aligned}$$

Therefore $F(V_n(x))$ is increasing in x. We have

$$W_n(x) \leq V_k(x+1), \text{ and } W_n(x) \leq F(V_n(x)) + c_s + c(x). \tag{9}$$

If $W_n(x+1) = V_n(x+2)$, then the first inequality in (9) proves that W_n is increasing. If $W_n(x+1) = F(V_n(x+1)) + c_s + c(x+1)$, then the second inequality in (9) proves that W_n is increasing because $F(V_n(x)) \leq F(V_n(x+1))$ and $c(x) \leq c(x+1)$.

We prove the convexity property of W_n. We may write

$$\begin{aligned}
W_n(x+1) + F(W_n(x+1)) + c(x+1) &\leq c(x+1) + V_n(x+2) + \sum_{h=0}^{x+1} q_{x+1,x+1-h} V_n(x+2-h) \\
&= c(x+1) + V_n(x+2) + F(V_n(x+2)) + \left(\frac{\gamma}{\gamma+\lambda}\right)^{x+2} (V_n(1) - V_n(0)),
\end{aligned} \tag{10}$$

and,

$$\begin{aligned}
W_n(x+1) + F(W_n(x+1)) + c(x+1) &\leq c(x+1) + F(V_n(x+1)) + c_s + c(x+1) \\
&+ \sum_{h=k+1}^{x+1} q_{x+1,x+1-h} V_n(x+2-h) + \sum_{h=0}^{k} q_{x+1,x+1-h} \left(F(V_n(x+1-h)) + c_s + c(x+1-h)\right),
\end{aligned} \tag{11}$$

for $2 \leq n \leq x$.

We distinguish two cases.

Case 1:

$$\begin{aligned}
W_n(x+2) + F(W_n(x)) + c(x) &= c(x) + V_n(x+3) + \sum_{h=0}^{x} q_{x,x-h} V_n(x+1-h) \\
&= c(x) + V_n(x+3) + F(V_n(x+1)) + \left(\frac{\gamma}{\gamma+\lambda}\right)^{x+1} (V_n(1) - V_n(0)) .
\end{aligned}$$

Since V_n is *igcv*, we have

$$V_n(x+2) + F(V_n(x+2)) + c(x+2) \leq V_n(x+3) + F(V_n(x+1)) + c(x+1).$$

Moreover, $\left(\frac{\gamma}{\gamma+\lambda}\right)^{x+2} (V_n(1) - V_n(0)) \leq \left(\frac{\gamma}{\gamma+\lambda}\right)^{x+1} (V_n(1) - V_n(0))$. Therefore,

$$\begin{aligned}
W_n(x+2) + F(W_n(x)) + c(x) &= c(x) + V_n(x+3) + F(V_n(x+1)) + \left(\frac{\gamma}{\gamma+\lambda}\right)^{x+1} (V_n(1) - V_n(0)) \\
&\geq V_n(x+2) + F(V_n(x+2)) + c(x+2) + c(x) - c(x+1) + \left(\frac{\gamma}{\gamma+\lambda}\right)^{x+2} (V_n(1) - V_n(0)) \\
&\geq V_n(x+2) + F(V_n(x+2)) + c(x+1) + \left(\frac{\gamma}{\gamma+\lambda}\right)^{x+2} (V_n(1) - V_n(0)) ,
\end{aligned}$$

since $c(x)$ is convex in x. Finally, Equation (10) proves that W_n is also *igcv*.

Case 2:

$$\begin{aligned}
W_n(x+2) + F(W_n(x)) + c(x) = c(x) + F(V_n(x+2)) + c_s + c(x+2) + \sum_{h=k+1}^{x} q_{x,x-h} V_n(x+1-h) \\
+ \sum_{h=0}^{k} q_{x,x-h} \left(F(V_n(x-h)) + c_s + c(x-h)\right) .
\end{aligned}$$

One may write

$$
\begin{aligned}
&F(V_n(x+1)) + c_s + 2c(x+1) + \sum_{h=k+1}^{x+1} q_{x+1,x+1-h} V_n(x+2-h) \\
&+ \sum_{h=0}^{k} q_{x+1,x+1-h} \left(F(V_n(x+1-h)) + c_s + c(x+1-h) \right) - F(V_n(x+2)) - c_s - c(x+2) - c(x) \\
&- \sum_{h=k+1}^{x} q_{x,x-h} V_n(x+1-h) - \sum_{h=0}^{k} q_{x,x-h} \left(F(V_n(x-h)) + c_s + c(x-h) \right) \\
&= F(V_n(x+1)) + c(x+1) - F(V_n(x+2)) - c(x+2) - c(x) \\
&+ \sum_{h=0}^{k} q_{x+1,x+1-h} \big(F(V_n(x+1-h)) + V_n(x+1-h) - F(V_n(x-h)) - V_n(x+2-h) \\
&\qquad\qquad + c(x+1-h) - c(x+2-h) \big) \\
&+ \sum_{h=0}^{x+1} q_{x+1,x+1-h} V_n(x+2-h) - \sum_{h=0}^{x} q_{x,x-h} V_n(x+1-h) \\
&= \sum_{h=0}^{k} q_{x+1,x+1-h} \big(F(V_n(x+1-h)) + V_n(x+1-h) - F(V_n(x-h)) - V_n(x+2-h) \\
&\qquad\qquad + c(x+1-h) - c(x+2-h) \big) \\
&- (V_n(1) - V_n(0)) \left(\frac{\gamma}{\gamma+\lambda} \right)^{x+1} \left(\frac{\lambda}{\gamma+\lambda} \right) + 2c(x+1) - c(x+2) - c(x) \leq 0,
\end{aligned}
$$

because V_n is *igcv* and $c(x)$ is increasing and convex. Hence, Equation (11) proves that W_n is also *igcv*.

We now assume that W_n and V_n are *igcv* and we prove that V_{n+1} is also *igcv*. We first prove that V_{n+1} is increasing in x. For $x = 0$, we have

$$
\begin{aligned}
V_{n+1}(1) - V_{n+1}(0) &= \gamma W_n(1) - \lambda W_n(0) + \mu \left(F(V_n(1)) + c(1) - V_n(0) \right) + (1 - \lambda - \mu)(V_n(1) - V_n(0)) \\
&+ (\lambda - \gamma) V_n(1) \\
&= \gamma(W_n(1) - W_n(0)) + \gamma(W_n(0) - V_n(0)) + \lambda(V_n(1) - W_n(0)) \\
&+ \mu u \left(V_n(1) - V_n(0) \right) + \mu c(1).
\end{aligned}
$$

The first term proportional with γ is positive since W_n is increasing in x, the second term in γ is positive because either $W_n(0) = V_n(1)$ and $W_n(0) - V_n(0) = V_n(1) - V_n(0) \geq 0$ or $W_n(0) = V_n(0) + c_s + c(0)$ and $W_n(0) - V_n(0) = c_s + c(0) \geq 0$, the term in λ is also positive because $W_n(0) = \min(V_n(1), V_n(0) + c_s + c(0)) \leq V_n(1)$, the other terms are also clearly positive. Therefore, $V_{n+1}(1) \geq V_{n+1}(0)$. For $x > 0$, we have

$$
\begin{aligned}
V_{n+1}(x+1) - V_{n+1}(x) &= \gamma(W_n(x+1) - W_n(x)) + \mu \left(F(V_n(x+1)) - F(V_n(x)) + c(x+1) - c(x) \right) \\
&+ (1 - \gamma - \mu)(V_n(x+1) - V_n(x)) \geq 0.
\end{aligned}
$$

Therefore V_{n+1} is increasing in x for $x \geq 0$.

We now prove that V_{n+1} is generally convex. For $x=0$, we may write

$$
\begin{aligned}
&V_{n+1}(2)+uV_{n+1}(0)-(1+u)V_{n+1}(1)-c(1)\\
&=\gamma(W_n(2)-(1+u)W_n(1)+uW_n(0)-c(1))\\
&+(1-\gamma-(1-u)\mu)(V_n(2)-(1+u)V_n(1)+uV_n(0)-c(1))\\
&+\mu(c(2)-2c(1))+uV(1)(\lambda+\mu(1-u))+uV_n(0)(1-\lambda-\mu-\mu(1-u)).
\end{aligned}
$$

The term proportional with γ is positive since W_n is igcv. The term proportional with $1-\gamma-(1-u)\mu$ is also positive since V_n is igcv. Since $c(0)=0$ and $c(x)$ is convex, $c(2)-2c(1)\geq 0$. The term proportional with $V_n(1)$ is positive. A necessary condition for the term proportional with $V_n(0)$ to be positive is $\mu(1-u)<\gamma$. This condition is satisfied if and only if $\gamma>\mu-\lambda$.

For $x>0$, we have

$$
\begin{aligned}
F(V_{n+1}(x)) &= \sum_{h=0}^{x} q_{x,x-h}V_{n+1}(x-h)\\
&=\gamma F(W_n(x))+(1-\gamma-\mu)F(V_n(x))+\mu\sum_{h=0}^{x-1}u(1-u)^h(F(V_n(x-h))+c(x-h))\\
&+\lambda(1-u)^xW_n(0)-\gamma(1-u)^xW_n(1)+(1-u)^x(1-\lambda)V_n(0)-(1-u)^x\lambda V_n(0),
\end{aligned}
$$

and

$$
\begin{aligned}
F(V_{n+1}(x+1)) &= \sum_{h=0}^{x+1} q_{x+1,x+1-h}V_{n+1}(x+1-h)\\
&=\gamma F(W_n(x+1))+(1-\gamma-\mu)F(V_n(x+1))+\mu\sum_{h=0}^{x-1}u(1-u)^h(F(V_n(x+1-h))\\
&\qquad +c(x+1-h))\\
&+\lambda(1-u)^{x+1}W_n(0)-\gamma(1-u)^{x+1}W_n(1)+(1-u)^{x+1}(1-\lambda)V_n(0)-(1-u)^{x+1}\lambda V_n(0)\\
&+\mu u(1-u)^x(uV_n(1)+(1-u)V_n(0)+c(1))
\end{aligned}
$$

So,

$$
\begin{aligned}
&V_{n+1}(x+2) - F(V_{n+1}(x)) + c(x) - V_{n+1}(x+1) - F(V_{n+1}(x+1)) - c(x+1) \qquad (12)\\
&\quad = \gamma(W_n(x+2) + F(W_n(x)) - W_n(x+1) - F(W_n(x+1)) + c(x) - c(x+1))\\
&\qquad + (1-\gamma-\mu)(V_n(x+2) + F(V_n(x)) - V_n(x+1) - F(V_n(x+1)) + c(x) - c(x+1))\\
&\qquad + \mu\sum_{h=0}^{x-1} u\,(1-u)^h\,(V_n(x+2-h) + F(V_n(x-h)) - V_n(x+1-h) - F(V_n(x+1-h))\\
&\qquad\qquad + c(x-h) - c(x+1-h))\\
&\qquad + \mu(c(x+2) + c(x) - 2c(x+1))\\
&\qquad + \mu u(1-u)^{x+1} V_n(1)\\
&\qquad + (\gamma - \mu(1-u))(1-u)^x V_n(0)\\
&\qquad + \mu u(1-u)^x(c(2) - uc(1))\\
&\qquad + u(1-u)^x(\lambda W_n(0) + \mu(1-u)(V_n(1) - V_n(0)) + \gamma(V_n(0) - W_n(1))).
\end{aligned}
$$

The first three terms of the expression are positive since V_n and W_n are *igcv*. The fourth term is positive since $c(x)$ is convex in x. The fifth term is also clearly positive. The sixth term is positive since $\gamma > \mu - \lambda$. We now consider the last two terms. Since $W_n(1) \leq V_n(1) + c_s + c(1)$, the last two terms are higher than

$$
\begin{aligned}
&\mu u(1-u)^x(c(2) - (1+u)c(1)) + u(1-u)^x(c(1)(\mu-\gamma) - \gamma c_s) \qquad (13)\\
&\qquad + u(1-u)^x(\lambda W_n(0) + (\mu-\lambda)(1-u)(V_n(1) - V_n(0))).
\end{aligned}
$$

We have $c(2) - (1+u)c(1) \geq 0$, since $c(x)$ is convex in x. Since $c(1) < c_s$ and $\gamma > \mu$, $c(1) < \frac{\gamma c_s}{\gamma-\mu}$. Finally, if $\mu > \lambda$, the term $(\mu-\lambda)(1-u)(V_n(1) - V_n(0))$ is positive. This finishes the proof of the Theorem. □

Proof of Corollary 1

We need to show that the constraints under which Theorem 1 holds are satisfied (or do not need to be satisfied) as n tends to infinity.

Since the stochastic process converges to the original one as γ tends to infinity, the constraints $\gamma > \mu$ and $c(1) < c_s$ are not restrictions of the result. As γ tends to infinity the time spent in the first waiting stage tends to zero, therefore $c(1)$ also tends to zero as γ tends to infinity.

The only real restriction of the result seems to be $\mu > \lambda$. This condition is only required to show that the last two terms of Equation (12) in the proof of Theorem 1 are together positive. We first showed that these terms are higher than or equal to Expression (13). In Expression (13), only the last term proportional with $\mu - \lambda$ may not be positive if $\mu < \lambda$. The term proportional with $\mu - \lambda$ is $u(1-u)^{x+1}(V_n(1) - V_n(0))$. As n tends to infinity $V_{n+1}(x) - V_n(x)$ converges to a finite limit (Puterman (1994)) for $x \geq 0$. As γ tends to infinity the cost of a container in stage 1 is zero, therefore $W_n(0) = V_n(1)$ as γ tends to infinity. We thus have $V_{n+1}(0) - V_n(0) = \lambda(W_n(0) - V_n(0)) = \lambda(V_n(1) - V_n(0))$. This proves that $V_n(1) - V_n(0)$ has a finite limit as n and γ tend to infinity. As γ tends to infinity $(1-u)^{x+1}$ tends to 1 and u tends to 0. Therefore $u(1-u)^{x+1}(V_n(1) - V_n(0))$ tends to 0 as γ tends to infinity and is negligible compared to the other terms in Expression (13). This proves that the condition $\mu > \lambda$ does not need to be satisfied to prove the optimality of a threshold policy in the long-run. □

Proof of Theorem 2

Stationary probabilities. Observing that

$$\left(\frac{\gamma}{\lambda+\gamma}\right)^x + \sum_{l=h}^{x-1} \left(\frac{\lambda}{\lambda+\gamma}\right)\left(\frac{\gamma}{\lambda+\gamma}\right)^l = \left(\frac{\gamma}{\lambda+\gamma}\right)^h, \tag{14}$$

we deduce that the cumulative transition rate from state x to states $0, 1, \cdots x-h$ is $\mu\left(\frac{\gamma}{\lambda+\gamma}\right)^h$, for $0 \leq h < x < n$; and that from state n to states $0, 1, \cdots n-h$ is $(\mu+\gamma)\left(\frac{\gamma}{\lambda+\gamma}\right)^h$, for $0 \leq h < n$. We now give the steady-state probability to be in state x, denoted by p_x, for $0 \leq x \leq n$.

Lemma 1 *We have*

$$p_0 = \frac{\gamma(\mu-\lambda)}{\gamma\mu + \lambda^2 - \lambda(\lambda+\gamma)\left(\frac{\lambda+\gamma}{\mu+\gamma}\right)^n},$$

$$p_x = \frac{\lambda}{\gamma}\left(\frac{\lambda+\gamma}{\mu+\gamma}\right)^x p_0, \textit{ for } 0 < x \leq n.$$

Proof. We first prove by induction on x that

$$p_{n-x} = \left(\frac{\mu+\gamma}{\lambda+\gamma}\right)^x p_n, \tag{15}$$

for $0 \leq x < n$. For $x = 0$, Equation (15) is straightforward. Assume now that Equation (15) is true for any rank r such that $0 \leq l \leq x$ and $0 \leq x < n-1$. Using Equation (14), we may write

$$\begin{aligned}
\gamma p_{n-(x+1)} &= \mu \sum_{l=1}^{x} \left(\frac{\gamma}{\lambda+\gamma}\right)^{x+1-l} p_{r-l} + (\mu+\gamma)\left(\frac{\gamma}{\gamma+\lambda}\right)^{x+1} p_n \\
&= \mu \sum_{l=1}^{x} \left(\frac{\gamma}{\lambda+\gamma}\right)^{x+1-l} \left(\frac{\mu+\gamma}{\lambda+\gamma}\right)^l p_n + (\mu+\gamma)\left(\frac{\gamma}{\lambda+\gamma}\right)^{x+1} p_n \\
&= (\mu+\gamma)\left(\frac{\gamma}{\lambda+\gamma}\right)^{x+1} \left(\frac{\mu+\gamma}{\gamma}\right)^x p_n.
\end{aligned}$$

This leads to $p_{n-(x+1)} = \left(\frac{\mu+\gamma}{\lambda+\gamma}\right)^{x+1} p_n$ and finishes the proof of Equation (15). Note now that for p_0, the transition rate from state 0 to 1 is λ instead of γ. Thus

$$p_0 = \frac{\gamma}{\lambda}\left(\frac{\mu+\gamma}{\lambda+\gamma}\right)^n p_n. \tag{16}$$

Since all probabilities sum up to one, we obtain p_0. This finishes the proof of the lemma. □

Probability of an empty system. If $\lambda \neq \mu$, we have $\frac{\lambda+\gamma}{\mu+\gamma} = \frac{\lambda+n/\tau}{\mu+n/\tau} = \left(1+\frac{\lambda}{n/\tau}\right)\frac{1}{1+\frac{\mu}{n/\tau}}$. As n tends to ∞, $\frac{1}{1+\frac{\mu}{n/\tau}} = 1-\frac{\mu}{n/\tau}+o(1/n)$. Thus as n tends to ∞, $\frac{\lambda+\gamma}{\mu+\gamma} = 1+\frac{\lambda-\mu}{n/\tau}+o(1/n)$. We also have as n tends to ∞, $\ln\left(\frac{\lambda+\gamma}{\mu+\gamma}\right) = \frac{\lambda-\mu}{n/\tau}+o(1/n)$. Then, $\lim\limits_{n\to\infty} n\ln\left(\frac{\lambda+\gamma}{\mu+\gamma}\right) = -\tau(\mu-\lambda)$, which implies $\lim_{n\to\infty}\left(\frac{\lambda+\gamma}{\mu+\gamma}\right)^n = e^{-\tau(\mu-\lambda)}$. If $\lambda = \mu$, we have $\frac{\lambda+\gamma}{\mu+\gamma} = 1$. Then, $\left(\frac{\lambda+\gamma}{\mu+\gamma}\right)^n = 1$, for $r \geq 1$ and we also have $\lim_{n\to\infty}\left(\frac{\lambda+\gamma}{\mu+\gamma}\right)^n = e^{-\tau(\mu-\lambda)}$. We therefore deduce the result of the proposition by letting n and γ go to infinity.

Other performance measures. We prove the equations for the case $\lambda \neq \mu$. The proofs for the case $\lambda = \mu$ follow from those of the case $\lambda \neq \mu$ by continuity. We consider the embedded Markov chain at matching initiation or rejection epochs. Matching initiations occur at μ-transitions from states $x > 0$. Rejection initiations only occur in state n with a γ-transition. The state probability just before a service initiation or a rejection is denoted by $\alpha(x)$. From flow conservation, we may write, $\alpha(x) = \frac{\mu p_x}{\lambda}$ for $0 < x < 0$, and $\alpha(n) = \frac{(\gamma+\mu)p_n}{\lambda}$ for $x = n$.

Proportion of containers sent back to the shipping line: It is given by $P_s = \lim\limits_{n\to\infty}\left(\frac{\gamma}{\lambda}p_n\right)$.

Expected time spent by container at the consignee: A matched container waits x γ-stages with probability $p_x\frac{\mu}{\lambda}$, for $0 < x \leq n$. We denote by T_S the time spent by a container which is sent to the shipper at the consignee. Averaging over all possibilities, we obtain

$$
\begin{aligned}
(1-P_s)\cdot E(T_S) &= \lim_{n\to\infty} \sum_{x=1}^{n}\frac{\mu}{\lambda}\frac{x}{\gamma}\frac{\lambda}{\gamma}\left(\frac{\lambda+\gamma}{\mu+\gamma}\right)^x p_0 \\
&= \lim_{n\to\infty} \frac{\mu}{\gamma^2}\sum_{x=1}^{n} x\left(\frac{\lambda+\gamma}{\mu+\gamma}\right)^x p_0 \\
&= \lim_{n\to\infty} \frac{\mu}{\gamma^2}\frac{\lambda+\gamma}{\mu+\gamma}\frac{-(n+1)(1-\frac{\lambda+\gamma}{\mu+\gamma})\left(\frac{\lambda+\gamma}{\mu+\gamma}\right)^n+1-\left(\frac{\lambda+\gamma}{\mu+\gamma}\right)^{n+1}}{\left(1-\frac{\lambda+\gamma}{\mu+\gamma}\right)^2}p_0.
\end{aligned}
$$

We therefore obtain $(1-P_s)\cdot E(T_S) = \frac{1-e^{-\tau(\mu-\lambda)}(1+\tau(\mu-\lambda))}{\mu(1-a)(1-ae^{-\tau(\mu-\lambda)})}$. Finally, $E(T) = (1-P_s)\cdot E(T_S) + P_s\cdot\tau$.

Waiting time distribution: A matched container can wait x γ-stages with probability $\frac{\mu}{\lambda}p_x$, for $0 < x \leq n$. For a container that matches from state $x > 0$, its waiting time is an Erlang random variable with x stages and a rate γ per stage. We thus have

$$
\begin{aligned}
(1-P_s)\cdot P(T_S > t) &= \lim_{n\to\infty} \sum_{x=1}^{n}\frac{\mu}{\lambda}\frac{\lambda}{\gamma}\left(\frac{\lambda+\gamma}{\mu+\gamma}\right)^x p_0\sum_{h=0}^{x-1}\frac{(\gamma t)^h}{h!}e^{-\gamma t} \\
&= \lim_{n\to\infty} \frac{\mu}{\lambda}\frac{\lambda}{\gamma}p_0 e^{-\gamma t}\sum_{x=1}^{n}\sum_{h=0}^{x-1}\frac{(\gamma t)^h}{h!}\left(\frac{\lambda+\gamma}{\mu+\gamma}\right)^x.
\end{aligned}
$$

We also may write

$$\sum_{x=1}^{n}\sum_{h=0}^{x-1}\frac{(\gamma t)^h}{h!}\left(\frac{\lambda+\gamma}{\mu+\gamma}\right)^x = \frac{\lambda+\gamma}{\mu+\gamma}\sum_{h=0}^{n-1}\frac{(\gamma t)^h}{h!}\sum_{x=h}^{n-1}\left(\frac{\lambda+\gamma}{\mu+\gamma}\right)^x$$
$$= \frac{\lambda+\gamma}{\mu+\gamma}\sum_{h=0}^{n-1}\frac{(\gamma t)^h}{h!}\left(\frac{\lambda+\gamma}{\mu+\gamma}\right)^h \frac{1-\left(\frac{\lambda+\gamma}{\mu+\gamma}\right)^{n-1-(h-1)}}{1-\frac{\lambda+\gamma}{\mu+\gamma}}$$
$$= \frac{\lambda+\gamma}{\mu-\lambda}\left(\sum_{h=0}^{n-1}\frac{(\gamma t)^h}{h!}\left(\frac{\lambda+\gamma}{\mu+\gamma}\right)^h - \sum_{h=0}^{n-1}\frac{(\gamma t)^h}{h!}\left(\frac{\lambda+\gamma}{\mu+\gamma}\right)^n\right).$$

Next, we get

$$\lim_{n\to\infty} e^{-\gamma t}\sum_{h=0}^{n-1}\frac{(\gamma t)^h}{h!}\left(\frac{\lambda+\gamma}{\mu+\gamma}\right)^n = e^{-\tau(\mu-\lambda)}, \text{ and } \lim_{n\to\infty} e^{-\gamma t}\sum_{h=0}^{n-1}\frac{(\gamma t)^h}{h!}\left(\frac{\lambda+\gamma}{\mu+\gamma}\right)^h = e^{-t(\mu-\lambda)}.$$

Therefore $(1-P_s)\cdot P(T_S>t) = \frac{e^{-t(\mu-\lambda)}-e^{-\tau(\mu-\lambda)}}{1-ae^{-\tau(\mu-\lambda)}}$ Using $P(T>t)=(1-P_s)\cdot P(T_S>t)+P_s$, we next obtain the result. □

Proof of Proposition 1

We compute the first and second order derivatives of the performance measures so as to evaluate their sign. Since all these measures are infinitely continuously derivable in τ for $\tau>0$, the strict convexity of the performance measures in the case $\lambda\neq\mu$ implies the convexity in the case $\lambda=\mu$. Moreover, note that the strict convexity for $\lambda=\mu$ holds also by following exactly the same approach as that for $\lambda\neq\mu$. In what follows, we focus on the case $\lambda\neq\mu$.

Let us first prove that $p_0(\infty)$ is decreasing and convex in τ. We may write

$$\frac{\partial p_0(\infty)}{\partial\tau} = -\frac{\lambda a(1-a)^2 e^{-\tau(\mu-\lambda)}}{\left(1-a^2e^{-\tau(\mu-\lambda)}\right)^2}<0.$$

Thus $p_0(\infty)$ is strictly decreasing in τ. We have

$$\frac{\partial^2 p_0(\infty)}{\partial\tau^2} = (1-a)^2\frac{2\left(a\lambda e^{-\tau(\mu-\lambda)}\right)^2+\left(1-a^2e^{-\tau(\mu-\lambda)}\right)\left(\lambda^2(1-a)e^{-\tau(\mu-\lambda)}\right)}{\left(1-a^2e^{-\tau(\mu-\lambda)}\right)^3},$$

or equivalently

$$\frac{\partial^2 p_0(\infty)}{\partial\tau^2} = \frac{a\lambda(1-a)^2e^{-\tau(\mu-\lambda)}\left(2a\lambda e^{-\tau(\mu-\lambda)}+\mu(1-a^2)+a\lambda\left(1-e^{-\tau(\mu-\lambda)}\right)\right)}{\left(1-a^2e^{-\tau(\mu-\lambda)}\right)^3}.$$

The sign of $\frac{\partial^2 p_0(\infty)}{\partial\tau^2}$ only depends on $2a\lambda e^{-\tau(\mu-\lambda)}+\mu(1-a^2)+a\lambda\left(1-e^{-\tau(\mu-\lambda)}\right)$ because the other parts of the expression are positive. We have

$$2a\lambda e^{-\tau(\mu-\lambda)}+\mu(1-a^2)+a\lambda\left(1-e^{-\tau(\mu-\lambda)}\right) = a\lambda e^{-\tau(\mu-\lambda)}+\mu>0.$$

Therefore, $\frac{\partial^2 p_0(\infty)}{\partial \tau^2} > 0$ and $p_0(\infty)$ is strictly decreasing and strictly convex in τ.

We next focus on the proportion of containers sent back to the shipping line, P_s. We have

$$\begin{aligned}\frac{\partial P_s}{\partial \tau} &= ae^{-\tau(\mu-\lambda)}\left(-(\mu-\lambda)p_0(\infty) + \frac{\partial p_0(\infty)}{\partial \tau}\right)\\ &= -ae^{-\tau(\mu-\lambda)}\frac{\mu(1-a)^2}{\left(1-a^2e^{-\tau(\mu-\lambda)}\right)^2} < 0.\end{aligned}$$

Therefore $\frac{\partial P_s}{\partial \tau} < 0$, and P_s is strictly decreasing in τ. Observe that

$$\frac{\partial P_s}{\partial \tau} = a\frac{\partial p_0(\infty)}{\partial \tau}.$$

Thus, $\frac{\partial P_s}{\partial \tau}$ is positively proportional to $\frac{\partial p_0(\infty)}{\partial \tau}$. Since $p_0(\infty)$ is strictly convex in τ, P_s is also strictly convex in τ.

We next focus on $E(T)$ and $E(N)$. We have

$$\begin{aligned}\frac{\partial E(T)}{\partial \tau} = &\frac{\mu}{(\mu-\lambda)^2}a\left[1+a+\frac{a^2}{1-a}\left(1-e^{-\tau(\mu-\lambda)}\right)\right]^{-2}\\ &\times\left[-\left(a\lambda e^{-\tau(\mu-\lambda)}\right)\left(1-(1+a\tau(\mu-\lambda))e^{-\tau(\mu-\lambda)}\right)\right.\\ &\left.\quad+\left(1+a+\frac{a^2}{1-a}\left(1-e^{-\tau(\mu-\lambda)}\right)\right)\left(e^{-\tau(\mu-\lambda)}\frac{(\mu-\lambda)^2}{\mu}(1+\tau\lambda)\right)\right].\end{aligned}$$

The sign of $\frac{\partial E(T)}{\partial \tau}$ depends on

$$\begin{aligned}&-\left(a\lambda e^{-\tau(\mu-\lambda)}\right)\left(1-(1+a\tau(\mu-\lambda))e^{-\tau(\mu-\lambda)}\right)\\ &+\left(1+a+\frac{a^2}{1-a}\left(1-e^{-\tau(\mu-\lambda)}\right)\right)\left(e^{-\tau(\mu-\lambda)}\frac{(\mu-\lambda)^2}{\mu}(1+\tau\lambda)\right).\end{aligned}$$

This expression is equivalent to

$$\begin{aligned}&(1+a)\left(e^{-\tau(\mu-\lambda)}\frac{(\mu-\lambda)^2}{\mu}(1+\tau\lambda)\right)\\ &+a\lambda e^{-\tau(\mu-\lambda)}\left(1-(1+a\tau(\mu-\lambda))e^{-\tau(\mu-\lambda)}+(1-a)(1+\lambda\tau)(1-e^{-\tau(\mu-\lambda)})\right).\end{aligned}$$

We have $(1+a)\left(e^{-\tau(\mu-\lambda)}\frac{(\mu-\lambda)^2}{\mu}(1+\tau\lambda)\right) > 0$ and $a\lambda e^{-\tau(\mu-\lambda)} > 0$. We next prove that $(1-a)(1+\lambda\tau)(1-e^{-\tau(\mu-\lambda)}) > 0$ and $1-(1+a\tau(\mu-\lambda))e^{-\tau(\mu-\lambda)} > 0$.

Case 1, $\lambda < \mu$: In this case, we have $-\tau(\mu-\lambda) < 0$ and $0 < e^{-\tau(\mu-\lambda)} < 1$. So, $0 < 1-e^{-\tau(\mu-\lambda)} < 1$. Moreover $1-a > 0$, thus, $(1-a)(1+\lambda\tau)(1-e^{-\tau(\mu-\lambda)} > 0$.

Case 2, $\lambda > \mu$: In this case we have $-\tau(\mu-\lambda) > 0$ and $e^{-\tau(\mu-\lambda)} > 1$. Then $1-e^{-\tau(\mu-\lambda)} < 0$. Moreover $1-a < 0$, so $(1-a)(1+\lambda\tau)(1-e^{-\tau(\mu-\lambda)}) > 0$.

We now consider the function $\Lambda(a)$ in a, defined $\Lambda(a) = 1-(1+a\tau\mu(1-a))e^{-\tau\mu(1-a)}$, for fixed values of μ and τ. We have $\Lambda'(a) = -e^{-\tau\mu(1-a)}(1-a)(2+a)\tau\mu$. If $a < 1$ then $\Lambda'(a) < 0$ and if $a > 1$ then $\Lambda'(a) > 0$. Thus Λ has a minimum at $a = 1$. Since $\lim_{a\to 1}\Lambda(a) = 0$, $\Lambda(a) > 0$ for $a \neq 1$ and $\frac{\partial E(T)}{\partial \tau} > 0$ for $a \neq 1$.

Therefore, $E(T)$ is strictly increasing in τ for $a \neq 1$.

It remains to prove that $P(T > t)$ is strictly increasing in τ for $t < \tau$. One may write

$$\frac{\partial P(T>\tau)}{\partial \tau} = a\mu e^{-\tau(\mu-\lambda)} \left[1+a+\frac{a}{1-a}\left(1-e^{-\tau(\mu-\lambda)}\right)\right]^{-2} \left(a\left(1+a+\frac{a}{1-a}\left(1-e^{-\tau(\mu-\lambda)}\right)\right) - \frac{e^{-t(\mu-\lambda)}-ae^{-\tau(\mu-\lambda)}}{1-a}a\right).$$

We have $\frac{e^{-t(\mu-\lambda)}-ae^{-\tau(\mu-\lambda)}}{1-a}a < 0$. So, $\frac{\partial P(T>\tau)}{\partial \tau} > 0$ and $P(T > \tau)$ is strictly increasing in τ for $t < \tau$. □

Proof of Proposition 2

We have

$$E(C) = c_s \frac{(1-a)e^{-\tau(\mu-\lambda)}}{1-ae^{-\tau(\mu-\lambda)}} + c_u \frac{1}{\mu} \frac{1-e^{-\tau(\mu-\lambda)}(1+a\tau(\mu-\lambda))}{(1-a)(1-ae^{-\tau(\mu-\lambda)})}.$$

We therefore obtain after some algebra,

$$\frac{\partial E(C)}{\partial \tau} = -\frac{e^{\mu\tau(1+a)}}{(ae^{a\mu\tau}-e^{\mu\tau})^2}\left(c_s\mu(1-a)^2 - c_u(1-2a+a(1-a)\mu\tau + a^2e^{-\mu(1-a)\tau})\right).$$

By deriving again the function in τ, $c_s\mu(1-a)^2 - c_u(1-2a+a(1-a)\mu\tau+a^2e^{-\mu(1-a)\tau})$ we can prove that this function is strictly decreasing in τ. Thus if $c_s\mu \leq c_u$, $\frac{\partial E(C)}{\partial \tau} \geq 0$ and the optimal threshold is $\tau = 0$. Otherwise, $\frac{\partial E(C)}{\partial \tau}$ is first negative, next positive. This prove that the expected cost has a unique minimum.

Proof of Proposition 3

Using proposition 2, we use the equality which allows to obtain τ in order to find $\frac{\partial \tau}{\partial c_s}$ and $\frac{\partial \tau}{\partial c_u}$. One may write

$$\begin{aligned}
\frac{\partial \tau}{\partial c_s} &= \frac{1-a}{c_u a(1-ae^{-\mu(1-a)\tau})} > 0, \\
\frac{\partial^2 \tau}{\partial c_s^2} &= -\frac{\partial \tau}{\partial c_s}\frac{a\mu(1-a)^2e^{-\mu(1-a)\tau}}{c_u a(1-ae^{-\mu(1-a)\tau})^2} < 0, \\
\frac{\partial \tau}{\partial c_u} &= -\frac{c_s(1-a)}{c_u^2 a(1-ae^{-\mu(1-a)\tau})} < 0, \\
\frac{\partial^2 \tau}{\partial c_u^2} &= c_s\frac{2c_u a(1-a)(1-ae^{-\mu(1-a)\tau}) + c_u^2a^2\mu(1-a)^2e^{-\mu(1-a)\tau}}{(c_u^2 a(1-ae^{-\mu(1-a)\tau}))^2} > 0.
\end{aligned}$$

Therefore, τ is increasing and concave in c_s and decreasing and convex in c_u. □