

# Performance analysis of a reflected fluid production/inventory model

Yonit Barron<sup>1</sup>

Received: 11 March 2015 / Accepted: 29 September 2015 / Published online: 12 October 2015  
© Springer-Verlag Berlin Heidelberg 2015

**Abstract** We study the performance of a reflected fluid production/inventory model operating in a stochastic environment that is modulated by a finite state continuous time Markov chain. The process alternates between ON and OFF periods. The ON period is switched to OFF when the content level reaches a predetermined level  $q$  and returns to ON when it drops to 0. The ON/OFF periods generate an alternative renewal process. Applying a matrix analytic approach, fluid flow techniques and martingales, we develop methods to obtain explicit formulas for the cost functionals (setup, holding, production and lost demand costs) in the discounted case and under the long-run average criterion. Numerical examples present the trade-off between the holding cost and the loss cost and show that the total cost appears to be a convex function of  $q$ .

**Keywords** Production–inventory model · Cost functionals · Fluid process · Martingale

## 1 Introduction

Inventory control has long focused on managing certain specific types and sources of uncertainty in the production and demand process. Manufacturing systems, operating in random environment, are implemented by a variety of manufacturing firms in the food-processing industry, such as sugar (Grunow et al. 2007), electronic computer industries and the pharmaceutical industry (see also Mohebbi 2006). In practice, the average production rate of these facilities is higher than the average demand. As a consequence, the production facility has to be switched off once in a while to pre-

---

✉ Yonit Barron  
ybarron@ariel.ac.il

<sup>1</sup> Industrial and Management Department, Ariel University, 40700 Ariel, Israel

vent inventory from growing without bound. In the special case of constant demand and production, the production–inventory system reduces to the classical economic production quantity (EPQ) model of Taft (1918). In the classic EPQ model every production cycle is composed of ON and OFF deterministic periods. There exists a predetermined level  $q$  such that the system is ON and the inventory level increases from level 0 up to level  $q$ . When level  $q$  is reached, the production is stopped and the inventory decreases down to 0. The time it takes from  $q$  to 0 is the OFF period.

In contrast with the assumption that production facilities can be switched off, many industries are characterized by high setup times and high setup costs where switching off the production is financially or operationally prohibitive. Thus, for both financial and operational reasons, it is critical to establish a proper production process early in the planning process; an overly high production rate results in high holding costs due to excess inventory, while a low production rate results in high penalty costs due to frequent stockouts and subsequent lost-sales.

In this paper we aim at filling some of this category. Specifically, we consider the heavy traffic version of a EPQ system in which neither production nor demand never stop (another interpretation to the latter assumption could be the case in which customers are allowed to return items to the system). Here, too, we assume two periods, ON and OFF; however, the production and demand rates differ during ON/OFF periods and are modulated by a continuous-time Markov chain (CTMC) which represents the environment. For example, the rates can change due to weather, economy, competition, seasonal promotion, customer status, forecasting, etc. We further assume that the average growth rate (which is the production rate minus the demand rate) is positive during ON periods and negative during OFF periods. Thus, each cycle is composed of two parts; the first part is the ON period which ends when the content level reaches a predetermined level  $q$  and the second part, the OFF period, is the time it takes from  $q$  until the content level reaches level 0; we assume that switching from one period to the other takes no time. Finally, we assume that backlogging is not allowed (i.e., demands that cannot be satisfied directly from the inventory are lost); hence, the content level is generated by reflection on the total production minus the total demand (satisfied or unsatisfied), with 0 acting as the reflecting barrier.

For background on stochastic EPQ models, Vickson (1986) studied a continuous review, single product stochastic cycling problem with demand modelled as a Brownian motion process. Recently, Wu and Chao (2013) extend this result to include a two-dimensional Brownian motion process. Berman and Perry (2001) present an ON/OFF model with a reflected Brownian motion. Boxma et al. (2001) consider a fluid system in which during the OFF times the buffer content increases as a piecewise linear process according to some semi-Markov process, and during the ON times, it decreases with a state-dependent rate. They derive the stationary distribution, steady state Laplace–Stieltjes transform and moments of the buffer content. Berman et al. (2007) study the long-run average performance of a fluid production/inventory model which alternates between ON periods and OFF periods and derive the pertinent reward functionals in a closed form.

Further examples on the ON/OFF process come from queueing literature (e.g., Kella and Whitt 1992; Boxma et al. 2005; Zhang and Zwart 2012), limited-capacity industries (e.g., Perry and Posner 2002), service systems, seasonal food products (i.e.,

ice cream, soup powder) and epidemics in which the quantity of interest is the number of susceptibles and the OFF period corresponds to mass preventive actions whenever the number of susceptibles exceed a specific number (Taylor 1968).

An example that the authors are familiar with comes from EL-AL, Israel Airline Ltd., ([www.elal.com](http://www.elal.com)). EL-AL has internal workshops that are responsible for the maintenance of the aircraft. Their activity is to repair the damaged aircraft spare parts. Huge cost and time associated with switching assembly lines off cause the workshops to operate continuously in a 24/7 manner. Due to the stochastic nature of the production (damaged components) and the demand (repaired items) processes, inventories (items waiting to be repaired) may reach undesirably high levels and can cause a shortage of repaired items essential for the airlines operations. Thus, each time a predetermined inventory level is reached, the workshop reduces the inventory level by increasing the rate of repair.

Another potential application of our study arises from customers which are allowed to return items to the buffer and negative inventory is not allowed. Inventory management with returning items have received a lot of attention in the literature; for example, Beltran and Krass (2002) experience with a major Canadian catalogue retailer that 30–40 % of inventory for many fashion items was due to returns, which made this source of “re-supply” very important. Recently, Shahrudin et al. (2015) study product returns and recovery management among six manufacturing firms in Malaysia. More examples of this line of research are Fleischmann et al. (2002) and Pinçer et al. (2008).

Our paper is also related to the literature encompasses a compound Poisson demand (e.g., Shi et al. 2014; Barron et al. 2014) or demand process that is a mixture of continuous demand and compound Poisson demand (e.g., Kok et al. 1984; Germs and Foreest 2014). For models based on Markovian decision process see e.g., Song and Zipkin (1993).

The underlying inventory process studied in this paper can also describe a variety of interpretations drawn from the contexts of perishable models. There are two types of perishable inventory models. The first family of models assumes that the quality of the items is slowly decreases over time; for example, a plant that produces ice-cream (see Boxma et al. 2014, 2015). The second family is models with obsolescences. Here, the items might perish at each period with some probability which is typically increasing over time. Examples of obsolescence models range from avionics and military sectors, high tech products, communications, construction equipment, medical devices, transportations and supply chain networks [we refer the readers to Tyler (2004), Song and Lau (2004), Shen and Willems (2014) for more examples].

In this paper we study four types of costs: (i) fixed setup cost; (ii) holding cost for the inventory; (iii) production cost and (iv) lost demand cost due to the reflection at level zero. Our objective is to obtain tractable formulas for the appropriate cost functionals. We focus on the discounted cost criterion using a discount factor  $\beta > 0$ ; however, we also deal with the long run average cost criterion. To the best of our knowledge, although the related literature is voluminous, our model seems to be more general and new and has been never investigated in the inventory literature; thus our research differs. Furthermore, while the afore-mentioned papers involve analytic derivations of the quantities of interest, we emphasize our study using a more probabilistic approach via

exit-time results and regenerative theory. this enables a simple derivation of quantities of interest and obtains easy-to-implement explicit formulas.

Our analysis is based on a combination of a certain martingale technique and an application of fluid flow theory. The martingale approach was introduced by [Asmussen and Kella \(2000\)](#) and was frequently used in the study of inventory models, see e.g., [Kella et al. \(2003\)](#), [Barron \(2015\)](#), [Barron et al. \(2014\)](#) and [Shi et al. \(2013\)](#) and the references given therein. Markov-modulated fluid flow models have been an active area of research in recent years; one of their main applications is to the modeling the traffic evolution in communication channels. [Silva and Latouche \(2005\)](#) consider fluid queue with a buffer of a finite capacity where the behavior of the background Markov process is allowed to change whenever the buffer is empty or when it is full. We use results of [Ramaswami \(1999\)](#) who initiated an unified matrix-analytic algorithmic approach to fluid flows, and this was followed by a series of papers by him, Ahn and others (see [Ramaswami 2006](#); [Ahn et al. 2007](#)).

Our analysis enables a simple derivation of quantities of interest and obtains easy-to-implement explicit formulas. These explicit formulas can then be used for an analysis of the dependence of the cost functionals on the system parameters, or for optimization purposes when some of these parameters (e.g., the costs, production rates and the storage capacity) are taken as decision variables. Through numerical examples we display the behavior of the costs and present the trade-off between the holding cost and the lost demand cost as a function of  $q$ . We show that the set up and holding costs are similar to those associated with the classical EPQ model. Finally, our observations show that the expected discounted total cost appears to be a convex function of  $q$ ; thus it is worthwhile for the controller to determine the optimal  $q$  in order to minimize the total cost.

The remainder of the paper is organized as follows. In Sect. 2 we present the mathematical description of the model and the cost functionals. The crucial tools of our analysis are introduced in Sect. 3. In Sect. 4 all the cost functionals for the discounted case derived in closed form. Finally, the long-run average analysis is given in Sect. 5. We also provide some numerical examples and insights.

## 2 Mathematical description of the model

We apply similar notations and definitions presented by [Barron \(2015\)](#). Let  $I(t)$  be the fluid level in the buffer at time  $t$ . The rate of change of the fluid level is modulated by a CTMC  $\{\mathcal{J}(t) : t \geq 0\}$  on a finite state space  $S = \{1, 2, \dots, n\}$  with a generator matrix  $Q = [Q_{ij}]$ . Let  $\pi = [\pi_1, \dots, \pi_n]$  be the limiting distribution of the  $\mathcal{J}(t)$  process, i.e.,  $\pi$  is the unique solution to  $\pi Q = 0$ ,  $\pi e = 1$  (where  $0$  is a row vector with all 0's and  $e$  is a column vector with all 1's). Let  $v = [v_1, v_2, \dots, v_n]$  be the initial probability vector of  $\mathcal{J}(t)$ . We assume  $I(0) = 0$ . The process  $I(t)$  can be partitioned into two parts,  $I^+(t)$  and  $I^-(t)$ . The first part of the cycle is the ON period  $\tau$  and this period ends whenever the content level reaches a predetermined level  $q$ . The second part of the cycle, the OFF period  $T$ , is the time it takes from  $\tau$  until the content level reaches level 0. Each period is characterized by stochastic inputs (production) and outputs (demand). However, we assume two sets of production rates  $\{p_1^1, \dots, p_n^1\} \subset (0, \infty)$

and  $\{p_1^2, \dots, p_n^2\} \subset (0, \infty)$  and two sets of demand rates  $\{d_1^1, \dots, d_n^1\} \subset (0, \infty)$  and  $\{d_1^2, \dots, d_n^2\} \subset (0, \infty)$ . The rate at which the inventory is filled at time  $t$  is determined by the current environmental state  $\mathcal{J}(t)$  and the period. During the ON period (OFF period) and as long as  $\mathcal{J}(t)$  is in state  $i$ , the production occurs continuously at rate  $p_i^1$  ( $p_i^2$ ), and there is a demand at rate  $d_i^1$  ( $d_i^2$ ). The growth rate is the difference of the production rate and the demand rate,  $r_i^1 = p_i^1 - d_i^1$  ( $r_i^2 = p_i^2 - d_i^2$ ). Note that  $r_i^1, r_i^2$   $i = 1, \dots, n$  may be either negative or positive. Let  $R_1 = \{r_1^1, \dots, r_n^1\}$  and  $R_2 = \{r_1^2, \dots, r_n^2\}$ . In our model we do not allow backlog; thus when the content level drops to level 0 (notice that this event can occur only during ON periods), it stays there as long as the environmental growth rate is negative and until the environmental state changes to some positive growth rate. Furthermore, the behavior of the process during each period is different; during the OFF period  $R_2$  satisfies

$$\sum_{i=1}^n \pi_i r_i^2 < 0, \tag{1}$$

and during the ON period,  $R_1$  satisfies

$$\sum_{i=1}^n \pi_i r_i^1 > 0. \tag{2}$$

A typical sample path of the inventory process is given in Fig. 1.

Note that, (1) is a necessary and sufficient condition for the stability of the OFF process (see Kulkarni and Yan 2007), while the ON process, with the absence of (2), is stable due to the reflection at level 0; however, (2) is assumed in order to avoid high lost demand. We assume that conditions (1) and (2) hold for the rest of this paper. Note that, due to the stochastic behavior of the process, the content level can exceed  $q$  during the OFF period.

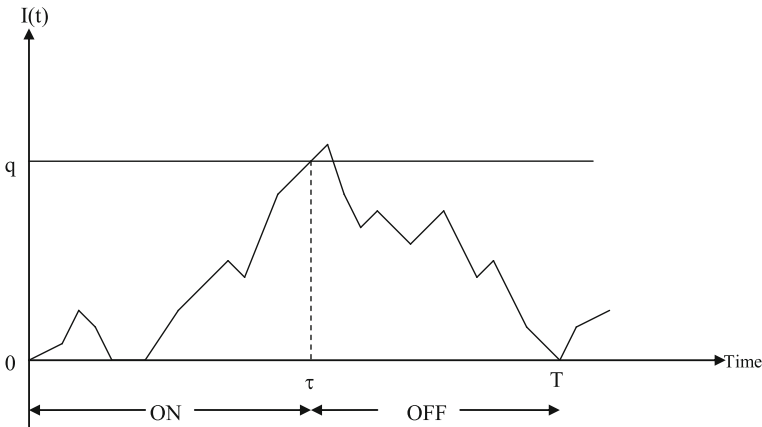


Fig. 1 A sample path of the inventory process  $I(t)$

*Remark 1* The model described above can be easily generalized by assuming arbitrary initial content level  $\gamma > 0$  (see e.g., [Berman et al. 2007](#)). Obviously, the long-run average cost is not affected; however, it affects the discounted cost. For simplicity, and since the case of  $\gamma > 0$  can be easily derived from our analysis, we assume  $\gamma = 0$ .

Clearly, the content level process is a semi-regenerative process which alternates between ON periods and OFF periods. Define the following stopping times:

$$\begin{aligned} T_0 &= 0 \\ T_n &= \inf \{t > \tau_n : I(t) = 0\} \quad n = 1, 2, \dots \\ \tau_n &= \inf \{t \geq T_{n-1} : I(t) = q\} \quad n = 1, 2, \dots \end{aligned}$$

$T_n, n = 1, 2, \dots$  are the times of switchings from OFF to ON and  $\tau_n, n = 1, 2, \dots$  are the time instants of switchings from ON to OFF. Thus, the content level process is a semi-regenerative process with  $T_n$ 's are semi-regenerative points. Define the  $n$ th cycle as the time elapsed between  $T_{n-1}$  and  $T_n, n = 1, 2, \dots$  and let  $C_n = T_n - T_{n-1}, n = 1, 2, \dots$  be the  $n$ th cycle length. Note that, if the ON periods are deleted from the sample path and the OFF periods are glued together we obtain a fluid EOQ model with refillings every time level 0 is reached.

For the rest of the paper, we use  $P(E)$  to denote the underlying probability measure (expectation), and  $P_i(E_i)$  denotes conditional probability (expectation) given initial state  $i \in S$ .  $\mathbf{E}$  and  $\mathbf{E}_i$  are the corresponding vector expectation and conditional vector expectation operators.  $\mathbb{E}, \mathbb{P}$  represent a matrix-valued expectation and probability operators, respectively. We denote by  $e_i$  a vector with the  $i$ th component equal to 1 and all the other components 0, by  $I$  the identity matrix (all of the appropriate sizes) and by  $\mathbf{1}_{\{A\}}$  the indicator of an event  $A$ .

## 2.1 The cost functionals

Let us now introduce the functionals representing the various expected discounted costs in our model, using the discount factor  $\beta > 0$ . We will use the fact that the inventory process is a semi renewal Markov process with  $T_n$ 's points as semi-regenerative points. For the rest of the paper let  $\tau = \tau_1, T = T_1, T' = T - \tau$  and  $C = C_1$  (so  $C = \tau + T'$ ). Note that  $I(t)$  is partitioned into two parts,  $I^+(t)$  (the ON period, the time until  $\tau$ ) and  $I^-(t)$  (the OFF period, the time that elapses from  $\tau$  until  $T$ ); furthermore, conditioning on the state at time  $\tau$  and the common background environmental process, the two process  $I^+(t)_{0 \leq t \leq \tau}$  and  $I^-(t)_{\tau \leq t \leq T}$  are independent.

In this paper we consider four costs: (a) the setup cost, (b) the holding cost of the inventory, (c) the production cost and (d) the unsatisfied demand cost. Our main objective is to develop techniques enabling us to determine all these costs in closed form under the discounted as well as under the long run average cost criterion. In the course of our derivations we will determine not only the cost functionals themselves but also various other transforms and probabilities describing the inventory level process that may be of independent interest.

- (a) *Set up cost* Let  $k_1$  be the setup cost to switch from OFF to ON (at time  $T$ ) and  $k_2$  be the setup cost to switch from ON to OFF (at time  $\tau$ ). The expected discounted set-up cost  $SC_\infty(\beta)$  is given by

$$SC_\infty(\beta) = \nu E \sum_{n=1}^{\infty} (k_1 \exp(-\beta T_n) + k_2 \exp(-\beta \tau_n)). \tag{3}$$

where  $\nu$  is the initial probability vector. Since  $I(t)$  is a semi-regenerative process,  $SC_\infty(\beta)$  can be expressed in terms of one cycle:

$$SC_\infty(\beta) = \nu \left( I - \mathbb{E} \left( e^{-\beta C} \right) \right)^{-1} \left( k_1 \mathbb{E} \left( e^{-\beta C} \right) + k_2 \mathbb{E} \left( e^{-\beta \tau} \right) \right) e. \tag{4}$$

We use  $\mathbb{E}(e^{-\beta C})$  as a shorthand notations for the  $(n \times n)$  matrix whose  $ij$ th component is given by

$$\left[ \mathbb{E} \left( e^{-\beta C} \right) \right]_{ij} = E \left( e^{-\beta C} \mathbf{1}_{\{\text{level 0 hit at time } C \text{ in phase } j\}} \mid \mathcal{J}(0) = i \right). \tag{5}$$

Similarly, the  $ij$ th component of the  $(n \times n)$  matrix  $\mathbb{E}(e^{-\beta \tau})$  is given by

$$\left[ \mathbb{E} \left( e^{-\beta \tau} \right) \right]_{ij} = E \left( e^{-\beta \tau} \mathbf{1}_{\{\text{level } q \text{ hit at time } \tau \text{ in phase } j\}} \mid \mathcal{J}(0) = i \right). \tag{6}$$

- (b) *Holding cost* The total expected discounted holding cost can be expressed by the functional

$$HC_\infty(\beta) = hE \left( \int_0^\infty e^{-\beta t} I(t) dt \right),$$

where  $hdt$  is the holding cost for a unit of stock during the time interval  $dt$ . Revoking the ergodic theorem for regenerative process, we can write  $HC_\infty(\beta)$  in terms of the first cycle and have

$$HC_\infty(\beta) = h\nu \left( I - \mathbb{E} \left( e^{-\beta C} \right) \right)^{-1} \left( \widehat{\mathbf{h}}_1(\beta) + \mathbb{E} \left( e^{-\beta \tau} \right) \widehat{\mathbf{h}}_2(\beta) \right),$$

where  $(n \times 1)$  vectors  $\widehat{\mathbf{h}}_1(\beta)$  and  $\widehat{\mathbf{h}}_2(\beta)$  are given by

$$\widehat{\mathbf{h}}_1(\beta) = \mathbf{E} \left( \int_0^\tau e^{-\beta t} I^+(t) dt \right), \quad \widehat{\mathbf{h}}_2(\beta) = \mathbf{E} \left( \int_0^{T'} e^{-\beta t} I^-(t + \tau) dt \right).$$

- (c) *Production cost* Let  $c$  be the production cost for one unit. Similarly, the expected discounted production cost  $PC_\infty(\beta)$  is given by

$$\begin{aligned} PC_\infty(\beta) &= cE \left( \int_0^\infty p_{\mathcal{J}(t)} e^{-\beta t} dt \right) \\ &= c\nu \left( I - \mathbb{E} \left( e^{-\beta C} \right) \right)^{-1} \left( \widehat{\mathbf{p}}_1(\beta) + \mathbb{E} \left( e^{-\beta \tau} \right) \widehat{\mathbf{p}}_2(\beta) \right), \end{aligned}$$

where  $\widehat{\mathbf{p}}_1(\beta)$  and  $\widehat{\mathbf{p}}_2(\beta)$  are  $(n \times 1)$  vectors represent the expected discounted production amount during ON and OFF periods, respectively.

$$\widehat{\mathbf{p}}_1(\beta) = \mathbf{E} \left( \int_0^\tau p_{\mathcal{J}(t)}^1 e^{-\beta t} dt \right), \quad \widehat{\mathbf{p}}_2(\beta) = \mathbf{E} \left( \int_0^{T'} p_{\mathcal{J}(t+\tau)}^2 e^{-\beta t} dt \right).$$

- (d) *Lost demand cost* In our mode, backlog is not allowed; any demand which cannot be satisfied immediately is lost. Clearly, there is no unsatisfied demand during the OFF period. During ON period, once level 0 is reached the process stays there until the environment changes to state with a positive growth rate. Assume the process hits 0 at state  $i$  (for some  $r_i^1 < 0$ ), the demand is lost with rate  $(-r_i^1)$  until the environmental state changes. Let  $w dt$  be the cost for a lost unit during a time interval of length  $dt$  ( $w > 0$ ). As a measure for the expected discounted lost demand cost one can use the functional

$$\begin{aligned} UC_\infty(\beta) &= -wE \left( \int_0^\infty e^{-\beta t} r_{\mathcal{J}(t)}^1 \mathbf{1}_{\{I^+(t)=0\}} dt \right) \\ &= -w\nu \left( I - \mathbb{E} \left( e^{-\beta C} \right) \right)^{-1} \mathbf{E} \left( \int_0^\tau e^{-\beta t} r_{\mathcal{J}(t)}^1 \mathbf{1}_{\{I^+(t)=0\}} dt \right) \\ &= w\nu \left( I - \mathbb{E} \left( e^{-\beta C} \right) \right)^{-1} \widehat{\mathbf{u}}(\beta). \end{aligned} \quad (7)$$

The  $(n \times 1)$  vector  $\widehat{\mathbf{u}}(\beta)$  is the expected discounted production loss during the first cycle. Notice that the lost demand can be written in terms of the local time process. Consider the process  $X(t) = \int_0^t r_{\mathcal{J}(u)}^1 du$  and let  $L(t) = -\min_{0 \leq u \leq t} X(u)$ . The process  $X(t)$  represents the inventory level at time  $t$  of the unrestricted process (the process without a barrier at level 0). The process  $L(t)$  is known as the *local time* and is a non-decreasing process that increases only whenever  $I^+(t) = 0$ . Therefore, an appropriate functional for the unsatisfied demand is

$$\widehat{\mathbf{u}}(\beta) = \mathbf{E} \left( \int_0^\tau e^{-\beta t} dL(t) \right). \quad (8)$$



A simple cost function for the entire system would be the sum, say  $TC(\beta)$ , of these four expected discounted costs:

$$TC_\infty(\beta) = SC_\infty(\beta) + PC_\infty(\beta) + HC_\infty(\beta) + UC_\infty(\beta). \tag{9}$$

The cost functionals  $\mathbb{E}(e^{-\beta C})$ ,  $\mathbb{E}(e^{-\beta \tau})$ ,  $\widehat{\mathbf{p}}_1(\beta)$ ,  $\widehat{\mathbf{p}}_2(\beta)$ ,  $\widehat{\mathbf{h}}_1(\beta)$ ,  $\widehat{\mathbf{h}}_2(\beta)$  and  $\widehat{\mathbf{u}}(\beta)$  will be derived in Sect. 4. The crucial tools of our analysis are introduced in the next section.

### 3 Preliminaries

For the determination of the cost functionals we use two tools: (a) an application of optional sampling theorem to a multi-dimensional martingale first presented by [Asmussen and Kella \(2000\)](#) and (b) the matrix-analytic approach and the theory of Markov-modulated fluid flows initiated by [Ramaswami \(2006\)](#). We summarize them shortly in order to make this paper self-contained.

#### 3.1 The multi-dimensional martingale

Let  $\{X(t), t \geq 0\}$  be a right continuous Markov modulated Lévy process with modulating process  $\{\mathcal{J}(t), t \geq 0\}$  which is a right continuous irreducible finite state space continuous time Markov chain. Let  $\{Y(t), t \geq 0\}$  be an adapted continuous process with a finite expected variation on finite intervals and let  $Z(t) = X(t) + Y(t)$ . [Asmussen and Kella \(2000\)](#) have shown, that for such a process, the matrix with elements  $E_i[e^{\alpha X(t)}; \mathcal{J}(t) = j]$  has the form of  $e^{tK(\alpha)}$  for some matrix  $K(\alpha)$ . Theorem 2.1 in [Asmussen and Kella \(2000\)](#) yields that under certain mild conditions on  $\{Z(t), t \geq 0\}$ , the multi-dimensional process

$$M(\alpha, t) = \int_0^t e^{\alpha Z(s)} \mathbf{1}_{\mathcal{J}(s)} ds K(\alpha) + e^{\alpha Z(0)} \mathbf{1}_{\mathcal{J}(0)} - e^{\alpha Z(t)} \mathbf{1}_{\mathcal{J}(t)} + \alpha \int_0^t e^{\alpha Z(s)} \mathbf{1}_{\mathcal{J}(s)} dY(s) \tag{10}$$

is a (row) vector valued zero mean martingale.

In our model, the inventory content level  $I(t)$  is a special case of  $X(t)$  and has piecewise linear sample paths with slope  $r_j^1$  during ON period and  $r_j^2$  during OFF period on intervals where  $\mathcal{J}(t) = j \in S$ ; Thus, some of the relevant functionals in this paper will be obtained by applying the optional stopping theorem (OST) (or Doob’s optional sampling theorem) to appropriate special cases of (10).

#### 3.2 The fluid inventory model

The conventional Markov-modulated fluid flow (MMFF) process is a fluid input-output process in which the rates of changes of fluid level are all linear and are governed by underlying Markov chain with infinitesimal generator  $Q$  and state space  $S$  ( $|S| = n$ ).

Let us partition the Markov chain phases as two sets  $\{S_1, S_2\}$  where  $S_1$  is the set of phases with increasing rates and  $S_2$  is the set of phases with decreasing rates ( $S = S_1 \cup S_2, |S_1| = n_1, |S_2| = n_2$ ). Comfortingly to  $S_1$  and  $S_2$ , the infinitesimal generator can be expressed as  $Q = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix}$ . During the sojourn time when the Markov phase is in  $i \in S_1$ , the fluid level increases with rate (slope)  $r_i > 0$  and decreases with slope  $|r_i|$  (for  $r_i < 0$ ) during  $i \in S_2$  (provided the fluid level is positive). We construct diagonal matrices  $U_j = \text{diag}\{|r_i|, i \in S_j\}, j = 1, 2$  and  $U = \text{diag}(U_1, U_2)$  from these rates.

Let  $\mathcal{F}(t)$  be the fluid level at time  $t$  and  $\mathcal{J}(t)$  the phase of the Markov chain at time  $t$ ; the two dimensional stochastic process  $\{\mathcal{F}(t), \mathcal{J}(t), t \geq 0\}$  is called the MMFF process. In the analysis of MMFF-related process, first passage times play important roles. Let  $\sigma(x) = \inf\{t > 0, \mathcal{F}(t) = x\}$  be the first passage time to level  $x$  and define the following Laplace–Stieltjes transforms (LST)

$$[\Psi(\beta)]_{ij} = E \left( e^{-\beta\sigma(0)}, \mathcal{J}(\sigma(0)) = j \mid \mathcal{F}(0) = 0, \mathcal{J}(0) = i \right) \quad i \in S_1, \quad j \in S_2. \tag{11}$$

$[\Psi(\beta)]_{ij}$  represents the LST of  $\sigma(0)$  restricted to the event that the fluid process hits level 0 at state  $j \in S_2$ , given  $\mathcal{F}(0) = 0$  and  $J(0) = i \in S_1$ . Ramaswami (2006, Appendix 1, p. 512) shows how to compute the matrix  $\Psi(\beta)$  and provides a good algorithm for this. Let  ${}_a^b\tau(x, y)$  (for  $x \geq 0, y \geq 0$ ) be the first passage time of  $\mathcal{F}$  from level  $x$  to level  $y$  avoiding a visit to the levels in  $[0, a] \cup [b, \infty)$  enroute (note that in the case of unlimited visit,  $a$  and  $b$  are omitted). We use the notation  ${}_a^b\widehat{f}(x, y, \beta)$  to denote the matrix of Laplace–Stieltjes transform of the joint distribution of the first passage time  ${}_a^b\tau(x, y)$  and the state of the phase process at each first passage time. Once we have computed  $\Psi(\beta)$ , the LST matrix  ${}_a^b\widehat{f}(x, y, \beta)$  is straightforward to evaluate. In the ‘‘Appendix’’, we present the detailed algorithm to compute  $\Psi(\beta)$  and the formulas for other LST matrices. For more details, we refer the readers to Ramaswami (2006).

Regarding our model, each ON/OFF period has one type of rates ( $R_1$  or  $R_2$ ); hence, given the state at switching epoch and the common environment, the ON/OFF periods are independent. Thus, we can analyze the inventory content level within each period independently using MMFF process. Specifically, we consider two fluid processes  $\mathcal{F}^+$  (with  $Q^+$ ) and  $\mathcal{F}^-$  (with  $Q^-$ ) corresponding to the ON and OFF periods, respectively; we also have the matrices  $U^+ = \text{diag}(U_1^+, U_2^+)$  and  $\Psi^+(\beta)$  ( $U^- = \text{diag}(U_1^-, U_2^-), \Psi^-(\beta)$ ) corresponding to the ON (OFF) process. We list in Table 1 LST matrices for hitting times, that we will use throughout our deriva-

**Table 1** LST of first passage times

LST	First passage time	Matrix size
${}_0\widehat{f}_{12}(x, 0, \beta)$	From $(x, S_1)$ to $(0, S_2)$ avoiding 0 in $\mathcal{F}$	$n_1 \times n_2$
${}_0\widehat{f}_{22}(x, 0, \beta)$	From $(x, S_2)$ to $(0, S_2)$ avoiding 0 in $\mathcal{F}$	$n_2 \times n_2$
${}_0\widehat{f}_{11}(0, x, \beta)$	From $(0, S_1)$ to $(x, S_1)$ avoiding 0 in $\mathcal{F}$	$n_1 \times n_1$
${}_x\Psi(\beta)$	From $(0, S_1)$ to $(0, S_2)$ avoiding $x$ in $\mathcal{F}$	$n_1 \times n_2$

tion. As mentioned before, these matrices are straightforward to evaluate once we have computed  $\Psi(\beta)$  and are given in Ramaswami (2006). Note that, for our model, each LST matrix should be derived for ON (marked as +) and OFF (marked as -) processes. For the rest of the paper we do not insert the marks + or - corresponding to ON/OFF processes; it should be clear from the context which of the marks applies.

*Remark 2* In the course of our derivations, we will apply the phase type distribution. A phase-type distribution with representation  $(\mathcal{E}, \alpha_0, \alpha, \mathbf{T})$  is the distribution of the time to absorption for a CTMC with a finite state space  $\mathcal{E}$  and one absorbing state.  $\alpha_0 \in [0, 1)$  is the probability of immediate absorption or the atom at 0; it is omitted if the distribution has no atom at 0.  $\alpha$  is the initial probability vector on the set of transient states and  $\mathbf{T}$  is the matrix of transition rates among the transient states. Let  $X$  be a phase type distributed, then

$$E(X) = \alpha \left(-\mathbf{T}^{-1}\right) e$$

$$M_X(\beta) = \alpha_0 + \alpha (\beta I - \mathbf{T})^{-1} T^o$$

where  $M_X(\beta)$  is the Laplace–Stieltjes transform of  $X$  and  $T^o = -\mathbf{T}e$ .

In this paper, for any matrix  $B$ , we shall denote its elements by  $(B)_{ij}$  or by  $[B]_{ij}$  and reserve the notation  $B_{ij}$  for the sub-matrix of  $B$  with row indices in  $S_i$  and column indices in  $S_j$ . Furthermore, for matrices  $A$  and  $B$  we use the shorthand notation  $(A, B)$  to denote the matrix obtained by stringing the matrix  $B$  after the matrix  $A$  (obviously,  $A$  and  $B$  have the same number of rows).

### 4 The discounted model

Let us now derive the seven functionals indicating the expected discounted costs in our model using a discount factor  $\beta > 0$ .

#### 4.1 Setup cost functionals

In order to obtain the setup costs we have to derive the  $(n \times n)$  matrices  $\mathbb{E}(e^{-\beta\tau})$  and  $\mathbb{E}(e^{-\beta C})$  [see (5), (6)].

##### 4.1.1 The matrix $\mathbb{E}(e^{-\beta\tau})$

Let  $\widehat{\mathbb{E}}(e^{-\beta\tau})$  be an  $(n \times n_1)$  matrix whose  $ij$ th component is the LST of the time to reach level  $q$  in phase  $j \in S_1$ , given initial phase  $i \in S$ . Since time  $\tau$  occurs only for states in  $S_1$ ,  $\mathbb{E}(e^{-\beta\tau})$  has the form of

$$\mathbb{E}(e^{-\beta\tau}) = \left(\widehat{\mathbb{E}}(e^{-\beta\tau}), \mathbf{0}_{(n \times n_2)}\right) \tag{12}$$

where  $\mathbf{0}_{a \times b}$  is a shorthand notation for an  $(a \times b)$  matrix with all entries equal to 0. For convenience, we will drop the size  $(a \times b)$  and apply  $\mathbf{0}$  in the cases that were previously mentioned (instead of  $\mathbf{0}_{a \times b}$ ).

Now, for the derivation of  $\widehat{\mathbb{E}}(e^{-\beta\tau})$  we introduce two similar LST matrices which differ only with their initial environment: (i) an  $(n_1 \times n_1)$  matrix  $\widehat{f}_{11}(0, q, \beta)$  whose  $ij$ th component  $(\widehat{f}_{11}(0, q, \beta))_{ij}$  represents the LST of the time until the content level hits  $q$  in state  $j \in S_1$ , given  $I(0) = 0$  and  $\mathcal{J}(0) = i \in S_1$  and (ii) an  $(n_2 \times n_1)$  matrix  $\widehat{f}_{21}(0, q, \beta)$  whose  $ij$ th component  $(\widehat{f}_{21}(0, q, \beta))_{ij}$  represents the LST of the time until the content level hits  $q$  in state  $j \in S_1$ , given  $I(0) = 0$  and  $\mathcal{J}(0) = i \in S_2$ .

**Proposition 4.1** *The matrix  $\widehat{f}_{11}(0, q, \beta)$  and  $\widehat{f}_{21}(0, q, \beta)$  are given by*

$$\widehat{f}_{11}(0, q, \beta) = {}_0\widehat{f}_{11}(0, q, \beta) + {}^q\Psi(\beta) (\beta I - Q_{22})^{-1} Q_{21} \widehat{f}_{11}(0, q, \beta) \tag{13}$$

$$\widehat{f}_{21}(0, q, \beta) = (\beta I - Q_{22})^{-1} Q_{21} \widehat{f}_{11}(0, q, \beta). \tag{14}$$

*Proof* Assume  $I(0) = 0, \mathcal{J}(0) \in S_1$ . There are two scenarios to reach level  $q > 0$ . In the first scenario,  $I(t)$  hits level  $q$  avoiding 0. Applying the fluid parameters and Table 1, the LST matrix of that time is  ${}_0\widehat{f}_{11}(0, q, \beta)$ . In the second scenario,  $I(t)$  drops to level 0 before reaching level  $q$  (with LST matrix  ${}^q\Psi(\beta)$ ). Once the process hits level 0, it stays there for a phase-type distributed time with transition matrix  $Q_{22}$ . Thus,  $(\beta I - Q_{22})^{-1} Q_{21}$  is the expected discounted time until exiting from level 0 with environmental in  $S_1$ . From that point, the process starts again from level 0. The proof of (14) is similar.  $\square$

The matrix  $\widehat{\mathbb{E}}(e^{-\beta\tau})$  is given by

$$\widehat{\mathbb{E}}(e^{-\beta\tau}) = \begin{pmatrix} \widehat{f}_{11}(0, q, \beta) \\ \widehat{f}_{21}(0, q, \beta) \end{pmatrix}. \tag{15}$$

Substituting (15) in (12) returns  $\mathbb{E}(e^{-\beta\tau})$ .

Notice that substituting  $\beta = 0$  in (12) and (15) returns the  $(n \times n)$  probability matrix

$$\mathbb{E}(\mathbf{1}_{\mathcal{J}(\tau)}) = \left( \begin{pmatrix} \widehat{f}_{11}(0, q, 0) \\ \widehat{f}_{21}(0, q, 0) \end{pmatrix}, \mathbf{0} \right). \tag{16}$$

The  $ij$ th component of  $\mathbb{E}(\mathbf{1}_{\mathcal{J}(\tau)})$  is the probability that the ON process hits level  $q$  at state  $j$  given  $\mathcal{J}(0) = i \in S$ .

#### 4.1.2 The matrix $\mathbb{E}(e^{-\beta C})$

Given the state at time  $\tau$ , the ON period and the OFF period are independent. Thus, regarding  $C = \tau + T'$  we obtain

$$\mathbb{E}(e^{-\beta C}) = \mathbb{E}(e^{-\beta\tau}) \mathbb{E}(e^{-\beta T'}).$$

$\mathbb{E}(e^{-\beta\tau})$  is given by (12). Let  $\widehat{\mathbb{E}}(e^{-\beta T'})$  be an  $(n \times n_2)$  matrix whose  $ij$ th components is the LST of the time to reach level 0 in phase  $j \in S_2$ , given  $\mathcal{J}(0) = i \in S, I(0) = q$ . Since time  $T'$  occurs only for states in  $S_2$ ,  $\mathbb{E}(e^{-\beta T'})$  has the form of

$$\mathbb{E}(e^{-\beta T'}) = (\mathbf{0}_{(n \times n_1)}, \widehat{\mathbb{E}}(e^{-\beta T'})) \tag{17}$$

It is easy to see that the matrix  $\widehat{\mathbb{E}}(e^{-\beta T'})$  is given by

$$\widehat{\mathbb{E}}(e^{-\beta T'}) = \begin{pmatrix} {}_0\widehat{f}_{12}(q, 0, \beta) \\ {}_0\widehat{f}_{22}(q, 0, \beta) \end{pmatrix} \tag{18}$$

where  ${}_0\widehat{f}_{12}(q, 0, \beta)$  and  ${}_0\widehat{f}_{22}(q, 0, \beta)$  are given in Table 1, with  $q$  replacing  $x$ . Notice that substituting  $\beta = 0$  in (17) and (18) returns the  $(n \times n)$  matrix

$$\mathbb{E}(\mathbf{1}_{J(T')}) = \left( \mathbf{0}, \begin{pmatrix} \widehat{f}_{12}(q, 0, 0) \\ \widehat{f}_{22}(q, 0, 0) \end{pmatrix} \right) \tag{19}$$

The  $ij$ th component of  $\mathbb{E}(\mathbf{1}_{J(T')})$  is the probability that the OFF process hits level 0 at state  $j$  given  $\mathcal{J}(\tau) = i$ .

Substituting (12) and (17) in (4) leads to the expected discounted set-up cost.

### 4.2 Holding cost

We start with the derivation of the  $(n \times 1)$  vector  $\widehat{\mathbf{h}}_1(\beta)$ , the expected discounted inventory level during the ON period. Recall that since we are now dealing with the ON period, for all the matrices in Table 1 we have to add the index + and use the corresponding matrices associated to  $\mathcal{F}^+(t)$ .

#### 4.2.1 $\widehat{\mathbf{h}}_1(\beta)$ Functional

The basic tool we use to derive  $\mathbf{E}(\int_0^\tau e^{-\beta t} I^+(t) dt)$  (and all other the required integrals, here and in the following sections) is optional sampling theory (OST) to the Asmussen–Kella multi-dimensional martingale, as introduced in Sect. 3.1. Consider the process

$$\begin{aligned} X(t) &= \int_0^t r_{\mathcal{J}(u)} du \quad 0 \leq t \leq \tau \\ X(0) &= 0 \end{aligned} \tag{20}$$

as defined in Sect. 2.1(d). Chapter XI, p. 311 of Asmussen (2003) yields that

$$E_i \left[ e^{\alpha X(t)}; \mathcal{J}(t) = j \right] = \left( e^{tK(\alpha)} \right)_{ij}$$

where

$$K(\alpha) = Q + \alpha U. \tag{21}$$

(specifically,  $K^+(\alpha) = Q^+ + \alpha U^+$  and  $Q^+, U^+$  are given in Sect. 3.2).

Recall that  $L(t) = -\min_{0 \leq u \leq t} X(u)$  is the local time and let  $Z(t) = X(t) + L(t)$ . It is not difficult to see that the latter process up to time  $\tau$ , i.e.,  $(Z(t))_{0 \leq t < \tau}$ , has the same distribution as  $(I(t))_{0 \leq t < \tau}$ . Finally, define  $Y(t) = L(t) - (\beta/\alpha)t$ , for arbitrary  $\beta \geq 0$  and  $\alpha < 0$ , and  $W(t) = X(t) + Y(t) = Z(t) - (\beta/\alpha)t$ . Since  $Y$  is adapted and has paths of finite expected total variation on bounded intervals, Theorem 2.1 of [Asmussen and Kella \(2000\)](#) yields that the process

$$\begin{aligned} M(\alpha, t) &= \left( M^1(\alpha, t), M^2(\alpha, t), \dots, M^n(\alpha, t) \right)^T \\ &= \int_0^t e^{\alpha W(u)} \mathbf{1}_{\mathcal{J}(u)} du K(\alpha) + e^{\alpha W(0)} \mathbf{1}_{\mathcal{J}(0)} - e^{\alpha W(t)} \mathbf{1}_{\mathcal{J}(t)} \\ &\quad + \alpha \left( \int_0^t e^{\alpha W(u)} \mathbf{1}_{\mathcal{J}(u)} dY(u) \right) \\ &= \int_0^t e^{\alpha Z(u) - \beta u} \mathbf{1}_{\mathcal{J}(u)} du (K(\alpha) - \beta I) + e^{\alpha Z(0)} \mathbf{1}_{\mathcal{J}(0)} - e^{\alpha Z(t) - \beta t} \mathbf{1}_{\mathcal{J}(t)} \\ &\quad + \alpha \left( \int_0^t e^{-\beta u} \mathbf{1}_{\mathcal{J}(u)} dL(u) \right) \end{aligned} \tag{22}$$

is an  $n$ -row vector-valued zero mean martingale. Here  $\mathbf{1}_{\mathcal{J}(t)}$  denotes the  $(1 \times n)$  vector  $\mathbf{1}_{\mathcal{J}(t)} = (\mathbf{1}_{\{\mathcal{J}(t)=1\}}, \dots, \mathbf{1}_{\{\mathcal{J}(t)=n\}})$ .

Recall that  $\mathbb{E}$  is the matrix-valued expectation operating on  $n$ -row vector random variables like  $M(\alpha, t)$  whose  $(i, j)$  component is the conditional expectation of the  $j$ th ( $j \in S$ ) entry given  $\mathcal{J}(0) = i \in S$ . For example,  $[\mathbb{E}(M(\alpha, t))]_{i,j} = [E(M^i(\alpha, t))]_j$ .

The OST yields  $\mathbb{E}(M(\alpha, \tau)) = \mathbb{E}(M(\alpha, 0)) = 0$ ,

$$\begin{aligned} \mathbb{E} \left( \int_0^\tau e^{\alpha Z(u) - \beta u} \mathbf{1}_{\mathcal{J}(u)} du \right) &= \left[ \mathbb{E} \left( e^{\alpha Z(\tau) - \beta \tau} \mathbf{1}_{\mathcal{J}(\tau)} \right) - \mathbb{E} \left( e^{\alpha Z(0)} \mathbf{1}_{\mathcal{J}(0)} \right) \right. \\ &\quad \left. - \alpha \mathbb{E} \left( \int_0^\tau e^{-\beta u} \mathbf{1}_{\mathcal{J}(u)} dL(u) \right) \right] (K(\alpha) - \beta I)^{-1}. \end{aligned} \tag{23}$$

Since  $\mathcal{J}(0) = i \in S$  and  $Z(0) = 0$  we obtain  $\mathbb{E}(e^{\alpha Z(0)} \mathbf{1}_{\mathcal{J}(0)}) = I$ . Applying  $Z(\tau) = q$  and the fluid method yields

$$\mathbb{E} \left( e^{\alpha Z(\tau) - \beta \tau} \mathbf{1}_{\mathcal{J}(\tau)} \right) = e^{\alpha q} \mathbb{E} \left( e^{-\beta \tau} \right) \tag{24}$$

where  $\mathbb{E}(e^{-\beta \tau})$  is given by (12).

To finish the computation of (23), we have to derive the  $(n \times n)$  matrix  $\mathbb{E}(\int_0^\tau e^{-\beta u} \mathbf{1}_{\mathcal{J}(u)} dL(u))$ , which is the expected discounted lost demand until time  $\tau$ .

First, we assume that the process starts at level 0 with environment in  $S_2$  and derive the expected lost demand until first exiting from 0. Notice that a loss occurs only during states in  $S_2$ . For that, we introduce the  $(n_2 \times n_2)$  matrix  $\Upsilon(\beta)$  whose  $ij$ th component,  $\Upsilon_{ij}(\beta)$ , is the expected discounted loss in state  $j \in S_2$  until exiting level 0, given that the process starts at level 0 with state  $i \in S_2$ .  $\Upsilon(\beta)$  is given by

$$\Upsilon(\beta) = -\mathbb{E} \left( \int_{t=0}^{\xi} \mathbf{r} e^{-\beta t} dt \right) + \mathbb{E} (e^{-\beta \xi}) \mathbb{P}_\Upsilon \Upsilon(\beta). \tag{25}$$

Note that once the inventory process hits 0 in state  $i \in S_2$ , it stays there for an exponential random time  $\xi_i$  with parameter  $-Q_{22}(i, i)$ . The matrix  $-\mathbb{E}(\int_{t=0}^{\xi} \mathbf{r} e^{-\beta t} dt)$  is an  $(n_2 \times n_2)$  diagonal matrix represents the expected discounted loss during time  $\xi_i$ ; its  $(ii)$ th component is given by  $-\frac{r_i}{\beta} \left( 1 + \frac{Q_{22}(i,i)}{\beta - Q_{22}(i,i)} \right)$ . From that point, with probability  $\frac{Q_{22}(i,j)}{-Q_{22}(i,i)}$ , the state changes to  $j \in S_2$ , and the expected discounted loss from state  $j$  is  $\Upsilon_{ji}(\beta)$ . Thus,  $\mathbb{E}(e^{-\beta \xi})$  is an  $(n_2 \times n_2)$  diagonal matrix whose  $(ii)$ th component is given by  $\frac{-Q_{22}(i,i)}{\beta - Q_{22}(i,i)}$  and  $\mathbb{P}_\Upsilon$  is an  $(n_2 \times n_2)$  probability matrix with  $(\mathbb{P}_\Upsilon)_{ij} = \frac{Q_{22}(i,j)}{-Q_{22}(i,i)}$  for  $i \neq j$  and 0 otherwise.

Now, we apply  $\Upsilon(\beta)$  to the derivation of the loss until time  $\tau$ . Since the lost demand occurs only for states in  $S_2$ , the matrix  $\mathbb{E}(\int_0^\tau e^{-\beta u} \mathbf{1}_{\mathcal{J}(u)} dL(u))$  has the form

$$\mathbb{E} \left( \int_0^\tau e^{-\beta u} \mathbf{1}_{\mathcal{J}(u)} dL(u) \right) = (\mathbf{0}, \widehat{\mathbf{L}}(\beta)) \tag{26}$$

where  $\widehat{\mathbf{L}}(\beta)$  is an  $(n \times n_2)$  matrix whose  $ij$ th component is the expected discounted loss in state  $j \in S_2$  until  $\tau$ , given  $\mathcal{J}(0) = i \in S$ . Let  $\widehat{\mathbf{L}}_1(\beta)$  be an  $(n_1 \times n_2)$  sub-matrix of  $\widehat{\mathbf{L}}(\beta)$  includes all rows corresponding to states in  $S_1$  and  $\widehat{\mathbf{L}}_2(\beta)$  be an  $(n_2 \times n_2)$  sub-matrix of  $\widehat{\mathbf{L}}(\beta)$  includes all rows corresponding to states in  $S_2$  such that

$$\widehat{\mathbf{L}}(\beta) = \begin{pmatrix} \widehat{\mathbf{L}}_1(\beta) \\ \widehat{\mathbf{L}}_2(\beta) \end{pmatrix}. \tag{27}$$

**Proposition 4.2**

$$\begin{aligned} \widehat{\mathbf{L}}_1(\beta) &= \left( I - {}^a \Psi(\beta) (\beta I - Q_{22})^{-1} Q_{21} \right)^{-1} {}^a \Psi(\beta) \Upsilon(\beta) \\ \widehat{\mathbf{L}}_2(\beta) &= \Upsilon(\beta) + (\beta I - Q_{22})^{-1} Q_{21} \widehat{\mathbf{L}}_1(\beta) \end{aligned} \tag{28}$$

*Proof* Assume  $\mathcal{J}(0) \in S_1$ .  $\widehat{\mathbf{L}}_1(\beta)$  satisfies the following equation

$$\widehat{\mathbf{L}}_1(\beta) = {}^q\Psi(\beta) \left( \Upsilon(\beta) + (\beta I - Q_{22})^{-1} Q_{21} \widehat{\mathbf{L}}_1(\beta) \right). \quad (29)$$

Lost demand occurs when the process drops to level 0 avoiding  $q$ . The LST of that event is  ${}^q\Psi(\beta)$ . From that point, by the proof of (25),  $\Upsilon(\beta)$  is the discounted lost demand. The term  $(\beta I - Q_{22})^{-1} Q_{21}$  is the discounted time until exiting level 0 with environmental state in  $S_1$  and starting again with LST  $\widehat{\mathbf{L}}_1(\beta)$ . Solving (29) for  $\widehat{\mathbf{L}}_1(\beta)$  yields (28). The matrix  $\widehat{\mathbf{L}}_2(\beta)$  has a similar proof.  $\square$

Substitute (26) and (24) in (23), multiply by  $e$  returns

$$\mathbf{E} \left( \int_0^\tau e^{\alpha Z(u) - \beta u} du \right) = (e^{\alpha q} \mathbb{E}(e^{-\beta \tau}) - I - \alpha (0_{n \times n_1}, \widehat{\mathbf{L}}(\beta))) (K(\alpha) - \beta I)^{-1} e. \quad (30)$$

Taking the derivation of (30) with respect to  $\alpha$  and setting  $\alpha = 0$  leads to

$$\begin{aligned} \widehat{\mathbf{h}}_1(\beta) &= \mathbf{E} \left( \int_0^\tau e^{-\beta u} I^+(u) du \right) = \mathbf{E} \left( \int_0^\tau e^{-\beta u} Z(u) du \right) \\ &= \frac{\partial}{\partial \alpha} \left[ (e^{\alpha q} \mathbb{E}(e^{-\beta \tau}) - I - \alpha (0_{n \times n_1}, \widehat{\mathbf{L}}(\beta))) (K(\alpha) - \beta I)^{-1} e \right] \Big|_{\alpha=0} \\ &= (q \mathbb{E}(e^{-\beta \tau}) - I - (0_{n \times n_1}, \widehat{\mathbf{L}}(\beta))) (Q - \beta I)^{-1} e - (\mathbb{E}(e^{-\beta \tau}) - I) \\ &\quad (Q - \beta I)^{-2} U e. \end{aligned} \quad (31)$$

which is the expected discounted inventory level during the ON period.

#### 4.2.2 $\widehat{\mathbf{h}}_2(\beta)$ Functional

In order to finish the holding cost we have to find the  $(n \times 1)$  vector  $\widehat{\mathbf{h}}_2(\beta) = \mathbf{E}(\int_0^{T'} e^{-\beta t} I^-(t) dt)$ . Recall that since we are now dealing with the OFF period, for all the matrices in Table 1 we have to add the index  $(-)$  and use the corresponding matrices associated to  $\mathcal{F}^-(t)$ . The OFF period starts at time  $\tau$ . We shift the time origin to  $\tau$  so that the OFF period starts at time 0. Consider  $\widetilde{X}(t)$  similar to (20) but with  $Q^-$  and  $U^-$  and  $\widetilde{X}(0) = q$ . Clearly, the latter process up to time  $T'$ , i.e.,  $(\widetilde{X}(t))_{0 \leq t < T'}$ , has the same distribution as  $(I(t))_{0 \leq t < T'}$ . Let  $\widetilde{W}(t) = \widetilde{X}(t) + \widetilde{Y}(t)$  with  $\widetilde{Y}(t) = -(\beta/\alpha)t$ , for arbitrary  $\beta \geq 0$  and  $\alpha < 0$ . Applying the optional sampling theory to the multi-



dimensional martingale (22) with  $T'$  replacing  $\tau$  we obtain

$$\mathbf{E} \left( \int_0^{T'} e^{\alpha \tilde{X}(s) - \beta s} ds \right) = \left[ \mathbb{E} \left( e^{\alpha \tilde{X}(T') - \beta T'} \mathbf{1}_{\mathcal{J}(T')} \right) - \mathbb{E} \left( e^{\alpha \tilde{X}(0)} \mathbf{1}_{\mathcal{J}(0)} \right) \right] \times (K(\alpha) - \beta I)^{-1} e. \tag{32}$$

Substituting  $\tilde{X}(0) = q$  and  $\tilde{X}(T') = 0$  returns

$$\begin{aligned} \mathbb{E} \left( e^{\alpha \tilde{X}(0)} \mathbf{1}_{\mathcal{J}(0)} \right) &= e^{\alpha q} I \\ \mathbb{E} \left( e^{\alpha \tilde{X}(T') - \beta T'} \mathbf{1}_{\mathcal{J}(T')} \right) &= \mathbb{E} \left( e^{-\beta T'} \right) \end{aligned} \tag{33}$$

where  $\mathbb{E}(e^{-\beta T'})$  is given by (17). Substituting (33) in (32), take the derivative with respect to  $\alpha$  and let  $\alpha = 0$  leads to

$$\begin{aligned} \hat{\mathbf{h}}_2(\beta) &= \mathbf{E} \left( \int_0^{T'} e^{-\beta s} I(s) ds \right) \\ &= \frac{\partial}{\partial \alpha} \left[ \left( \left( \mathbf{0}, \begin{pmatrix} \hat{f}_{12}(q, 0, 0) \\ \hat{f}_{22}(q, 0, 0) \end{pmatrix} \right) - e^{\alpha q} I \right) (K(\alpha) - \beta I)^{-1} e \right] \Big|_{\alpha=0} \\ &= -qI(Q - \beta I)^{-1} e - \left( \left( \mathbf{0}, \begin{pmatrix} \hat{f}_{12}(q, 0, 0) \\ \hat{f}_{22}(q, 0, 0) \end{pmatrix} \right) - I \right) (Q - \beta I)^{-2} Ue. \end{aligned} \tag{34}$$

### 4.3 Production cost

The expected discounted production vector costs  $\hat{\mathbf{p}}_1(\beta) = \mathbf{E}(\int_0^\tau e^{-\beta s} p_{\mathcal{J}(s)}^1 ds)$  and  $\hat{\mathbf{p}}_2(\beta) = \mathbf{E}(\int_0^{T'} e^{-\beta s} p_{\mathcal{J}(s)}^2 ds)$  (of order  $(n \times 1)$ ) can be easily derived as follows. For the derivation of  $\hat{\mathbf{p}}_1(\beta)$  (ON period), let  $\mathbf{C}_1^+, \mathbf{C}_2^+$  and  $\mathbf{C}^+$  be diagonal matrices as follows:

$$\begin{aligned} \mathbf{C}_j^+ &= \text{diag} \left\{ p_i^1, i \in S_j \right\}, \quad j = 1, 2. \\ \mathbf{C}^+ &= \text{diag} \left( \mathbf{C}_1^+, \mathbf{C}_2^+ \right). \end{aligned} \tag{35}$$

Let  $\mathbb{E}(\int_0^\tau e^{-\beta s} \mathbf{1}_{\mathcal{J}(s)} ds)$  be an  $(n \times n)$  matrix whose  $(i, j)$  component represents the expected discounted time in phase  $j \in S$  given  $I(0) = i \in S$ . Clearly,  $\hat{\mathbf{p}}_1(\beta)$  is given by

$$\hat{\mathbf{p}}_1(\beta) = \mathbb{E} \left( \int_0^\tau e^{-\beta s} \mathbf{1}_{\mathcal{J}(s)} ds \right) \mathbf{C}^+ e,$$

Applying (23) to get

$$\mathbb{E} \left( \int_0^\tau e^{\alpha Z(s) - \beta s} \mathbf{1}_{\mathcal{J}(s)} ds \right) = (e^{\alpha q} \mathbb{E}(e^{-\beta \tau}) - I - \alpha (\mathbf{0}, \widehat{\mathbf{L}}(\beta))) (K(\alpha) - \beta I)^{-1}. \quad (36)$$

Substituting  $\alpha = 0$  yields

$$\mathbb{E} \left( \int_0^\tau e^{-\beta s} \mathbf{1}_{\mathcal{J}(s)} ds \right) = (\mathbb{E}(e^{-\beta \tau}) - I) ((0) - \beta I)^{-1}. \quad (37)$$

Now multiplying (37) by the vector  $\mathbf{C}^+ e$  returns the expected discounted production cost,  $\widehat{\mathbf{p}}_1(\beta)$ , for the ON process.

Regarding  $\widehat{\mathbf{p}}_2(\beta)$  and similar to above, we apply (32) to get the  $(n \times n)$  matrix

$$\begin{aligned} \mathbb{E} \left( \int_0^{T'} e^{\alpha \tilde{X}(s) - \beta s} \mathbf{1}_{\mathcal{J}(s)} ds \right) &= \left[ \mathbb{E} \left( e^{\alpha \tilde{X}(T') - \beta T'} \mathbf{1}_{\mathcal{J}(T')} \right) - \mathbb{E} \left( e^{\alpha \tilde{X}(0)} \mathbf{1}_{\mathcal{J}(0)} \right) \right] \\ &\quad (K(\alpha) - \beta I)^{-1} \\ &= \left( \mathbb{E} \left( e^{-\beta T'} \right) - e^{\alpha q} I \right) (K(\alpha) - \beta I)^{-1}. \end{aligned} \quad (38)$$

Now, define  $\mathbf{C}^-, \mathbf{C}_j^-, j = 1, 2$  as in (35) with  $p_i^2$  replacing  $p_i^1$ . Substitute  $\alpha = 0$  in (38) and multiply by the matrix  $\mathbf{C}^-$  we obtain the expected discounted production rate for the OFF process

$$\widehat{\mathbf{p}}_2(\beta) = \mathbb{E} \left( \int_0^{T'} e^{-\beta s} \mathbf{1}_{\mathcal{J}(s)} ds \right) \mathbf{C}^- e = \left( \mathbb{E} \left( e^{-\beta T'} \right) - I \right) (K(0) - \beta I)^{-1} \mathbf{C}^- e.$$

#### 4.4 Loss cost

To end our discounted analysis, we need to derive the  $(n \times 1)$  expected discounted production loss vector  $\widehat{\mathbf{u}}(\beta) = \mathbf{E} \left( \int_0^\tau e^{-\beta t} dL(t) \right)$ .

**Corollary 4.1** *It is easy to see that  $\widehat{\mathbf{u}}(\beta) = \widehat{\mathbf{L}}(\beta)e$ .*

We summarize the main computation steps of the performance measures for the discounted case:

1. Compute  $\Psi(\beta)$  and the LST matrices (see Table 1).
2. Obtain  $\widehat{f}_{11}(0, q, \beta)$  and  $\widehat{f}_{21}(0, q, \beta)$  using (13) and (14).

3. Compute  $\mathbb{E}(e^{-\beta\tau})$ ,  $\widehat{\mathbb{E}}(e^{-\beta\tau})$  and  $\mathbb{E}(e^{-\beta C})$  [Eqs. (12)–(18)] which leads to  $SC_\infty(\beta)$ .
4. Applying (31) and (34) returns  $\widehat{\mathbf{h}}_1(\beta)$  and  $\widehat{\mathbf{h}}_2(\beta)$  which leads to  $HC_\infty(\beta)$ .
5. Applying (37) and (38) to obtain  $\widehat{\mathbf{p}}_1(\beta)$  and  $\widehat{\mathbf{p}}_2(\beta)$ , and thus  $PC_\infty(\beta)$ .
6. Finally, Corollary 4.1 returns  $\widehat{\mathbf{u}}(\beta)$  which leads to  $UC_\infty(\beta)$ .

*Example 1* We consider a Markov chain with  $n = 4$  states with the following generator matrix

$$Q = \begin{pmatrix} -7 & 1.5 & 2.5 & 3 \\ 3 & -9 & 2 & 4 \\ 2 & 4 & -9 & 3 \\ 4 & 3 & 2 & -9 \end{pmatrix}.$$

As initial probability vector we take  $\nu = (0.2, 0.3, 0.35, 0.15)$ . We take the following basic values for the parameters:

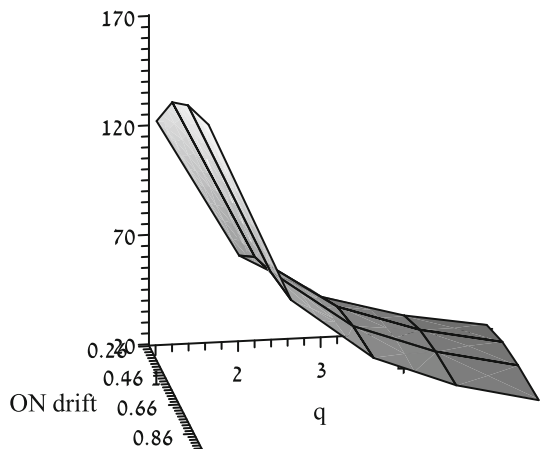
$$\beta = 0.01, \quad k_1 = 5, \quad k_2 = 5, \quad c = 3, \quad h = 1, \quad w = 1.$$

Let  $\gamma_1$  and  $\gamma_2$  be the average growth rates (in short, drifts) of the ON and OFF periods, respectively, i.e.,

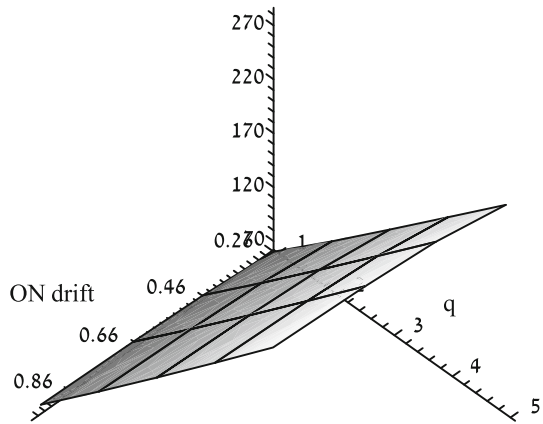
$$\gamma_1 = \sum_{i=1}^n \pi_i r_i^1 > 0, \quad \gamma_2 = \sum_{i=1}^n \pi_i r_i^2 < 0$$

We assume  $\gamma_2 = -0.214$ . Let  $\gamma_1$  vary in  $\{0.207, 0.48, 0.69, 0.96\}$  (by changing the production rates during the ON period) and  $q$  vary in  $\{1, 2, 3, 4, 5\}$ . We determine the expected discounted setup, holding, loss and total cost. We do not reproduce lengthy tables of calculated values (which could be easily done, but is not very illuminating); instead we present our results in terms of graphical displays. Figures 2, 3 and 4 shows

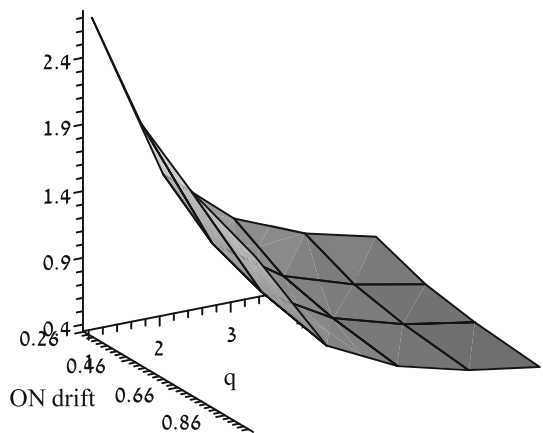
**Fig. 2** The expected discounted setup cost,  $SC_\infty(\beta)$



**Fig. 3** The expected discounted holding cost,  $HC_\infty(\beta)$



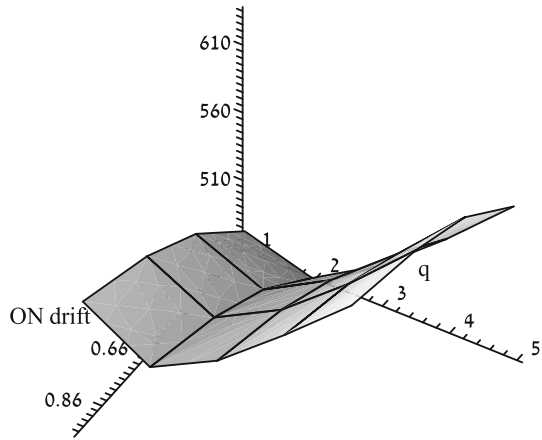
**Fig. 4** The expected discounted loss cost,  $UC_\infty(\beta)$



$SC_\infty(\beta)$ ,  $HC_\infty(\beta)$ ,  $UC_\infty(\beta)$  and  $TC_\infty(\beta)$  as functions of  $\gamma_1$  and  $q$ , respectively. We see that  $SC_\infty(\beta)$  decreases as a function of  $q$  and almost unaffected by  $\gamma_1$ . Figure 3 shows that  $HC_\infty(\beta)$  increases as a function of both  $\gamma_1$  and  $q$ ; however,  $UC_\infty(\beta)$  has an opposite behavior, since increasing production rates or increasing  $q$  decrease the number of drops to level 0 and cause less loss. In Fig. 5 we see that the total cost increases as a function of  $\gamma_1$  and appears to be a convex function of  $q$ ; thus we can conclude that for each  $\gamma_1$  there is an optimal  $q$  which minimize the total cost. Notice that, the behavior of  $SC_\infty(\beta)$ ,  $HC_\infty(\beta)$  and  $TC_\infty(\beta)$  as a function of  $q$  is similar to those associated with the classical EPQ model.

Furthermore, since we see that the expected discounted total cost appears to be a convex function of  $q$ , we have been carried out additional numerical checkings and present in Table 2 the optimal value  $q^*$  and  $TC_\infty^*(\beta)$  for  $\gamma_1$  vary in  $\{0.207, 0.48, 0.69, 0.96\}$ . We can see that the optimal value  $q^*$  and the optimal discounted total cost  $TC_\infty^*(\beta)$  increase in  $\gamma_1$ .

**Fig. 5** The expected discounted total cost,  $TC_\infty(\beta)$



**Table 2**  $q^*$  and  $TC_\infty^*(\beta)$  for  $\gamma_1$  vary in  $\{0.207, 0.48, 0.69, 0.96\}$

$\gamma_1$	$q^*$	$TC_\infty^*(\beta)$
0.207	1.6	457.96
0.48	1.69	493.006
0.69	1.74	511.216
0.96	1.78	524.768

### 5 Long run average analysis

Let  $TC(t)$  be the total cost until time  $t$ . We are interested in the long run average total cost per time unit:

$$\overline{TC} = \lim_{t \rightarrow \infty} \frac{TC(t)}{t}.$$

Regarding  $T_n$ , let  $\mathcal{J}_n = \mathcal{J}(T_n), n = 0, 1, \dots$  be the environmental state at  $T_n$  (at the beginning of each cycle). Notice that  $(\mathcal{J}_n, T_n)$  is a Markov renewal process, and the content level process  $I(t)$  is a semi-regenerative process with respect to  $(\mathcal{J}_n, T_n)$ .  $T_n$ 's are semi-regenerative points of the process. Thus, the process  $\mathcal{J}_n, n = 0, 1, 2, \dots$  is an irreducible positive recurrent Markov chain; we denote its transition probability matrix by  $\mathbb{P}^*$ . Let  $\pi^*$  be the corresponding stationary probability vector.  $\pi^*$  is determined by the following equations:

$$\begin{aligned} \pi^* &= \pi^* \mathbb{E}(\mathbf{1}_{\mathcal{J}(\tau)}) \mathbb{E}(\mathbf{1}_{\mathcal{J}(T')}) \\ \pi^* e &= 1, \end{aligned} \tag{39}$$

where  $\mathbb{E}(\mathbf{1}_{\mathcal{J}(\tau)})$  and  $\mathbb{E}(\mathbf{1}_{\mathcal{J}(T')})$  are given in (16) and (19), respectively.

The long run average cost  $\overline{TC}$  can be derived from  $TC_\infty(\beta)$  by a well-known procedure

$$\overline{TC} = \lim_{\beta \rightarrow 0} \beta TC(\beta) = \overline{SC} + \overline{HC} + \overline{PC} + \overline{UC}$$

where on the right side we have the long-run average costs of the set-up, holding stock, production and unsatisfied demand, respectively.

Let  $N(t)$  be the number of cycles until time  $t$ , i.e.,  $N(t) = \sup\{n \geq 0 \mid T_0 + \dots + T_n \leq t\}, t \geq 0$ . Following [Asmussen \(2003\)](#), Chapter VI, Theorem 3.1 and Chapter VII, Proposition 5.2, the four long-run quantities can be given in terms of the first-cycle functionals as follows:

$$\begin{aligned} \overline{SC} &= \lim_{t \rightarrow \infty} \frac{N(t)(k_1 + k_2)}{t} = \frac{k_1 + k_2}{E(C)}. \\ \overline{HC} &= \lim_{t \rightarrow \infty} \frac{hE\left(\int_0^t I(s)ds\right)}{t} = \frac{h \sum_i \pi_i^* E_i\left(\int_0^C I(t)dt\right)}{E(C)} \\ &= \frac{h\pi^* (\widehat{\mathbf{h}}_1(0) + \mathbb{E}(\mathbf{1}_{\mathcal{J}(\tau)})\widehat{\mathbf{h}}_2(0))}{E(C)}. \\ \overline{PC} &= \lim_{t \rightarrow \infty} \frac{cE\left(\int_0^t p_{\mathcal{J}(s)}ds\right)}{t} = \frac{c \sum_i \pi_i^* E_i\left(\int_0^C p_{\mathcal{J}(t)}dt\right)}{E(C)} \\ &= \frac{c\pi^* (\widehat{\mathbf{p}}_1(0) + \mathbb{E}(\mathbf{1}_{\mathcal{J}(\tau)})\widehat{\mathbf{p}}_2(0))}{E(C)}. \\ \overline{UC} &= \lim_{t \rightarrow \infty} \frac{wE\left(\int_0^t dL(s)\right)}{t} = \frac{w\pi^*\widehat{\mathbf{u}}(0)}{t} = \frac{w\pi^*\mathbf{E}(L(\tau))}{E(C)}. \end{aligned}$$

Note that  $E(C)$  denotes the steady-state expected value of  $C$ . Clearly,  $E(C) = \pi^*\mathbf{E}(C)$  where  $\mathbf{E}(C)$  is an  $(n \times 1)$  vector whose  $i$ th component is  $E_i(C) = E(C \mid \mathcal{J}_0 = i \in S)$ .

To apply the above equations we have to compute  $\widehat{\mathbf{h}}_1(0), \widehat{\mathbf{h}}_2(0), \widehat{\mathbf{p}}_1(0), \widehat{\mathbf{p}}_2(0), \mathbf{E}(L(\tau)), \mathbf{E}(C)$  and  $\pi^*$ .

### 5.1 Long-run average inventory level

In order to derive  $\widehat{\mathbf{h}}_1(0)$  set  $\beta = 0$  in (23) leads to

$$\begin{aligned} \mathbf{E}\left(\int_0^\tau e^{\alpha Z(u)} du\right) &= \left[ \mathbb{E}\left(e^{\alpha Z(\tau)} \mathbf{1}_{\mathcal{J}(\tau)}\right) - \mathbb{E}\left(e^{\alpha Z(0)} \mathbf{1}_{\mathcal{J}(0)}\right) \right. \\ &\quad \left. - \alpha \mathbb{E}\left(\int_0^\tau \mathbf{1}_{\mathcal{J}(u)} dL(u)\right) \right] K(\alpha)^{-1} e. \end{aligned} \tag{40}$$

Applying  $Z(\tau) = q$  leads to the  $(n \times n)$  matrix

$$\mathbb{E} \left( e^{\alpha Z(\tau)} \mathbf{1}_{J(\tau)} \right) = e^{\alpha q} \mathbb{E} \left( \mathbf{1}_{J(\tau)} \right).$$

Clearly  $\mathbb{E}(e^{\alpha Z(0)} \mathbf{1}_{J(0)}) = I$ . For the final term, set  $\beta = 0$  in (26)–(28). Notice that once the process drops to level 0, it stays there for a phase type distributed random variable with transition rate matrix  $Q_{22}$ . Thus

$$\Upsilon(0) = (-Q_{22})^{-1}.$$

(for a proof, see [Latouche and Ramaswami 1999](#), Theorem 2.4.3). Now take the derivation of (40) with respect to  $\alpha$  and set  $\alpha = 0$ . However this is not possible as  $K(\alpha)$  is singular at  $\alpha = 0$ . Instead we have to use (23) to conclude that

$$\frac{\partial}{\partial \alpha} \left[ (e^{\alpha q} \mathbb{E}(\mathbf{1}_{\mathcal{J}(\tau)}) - I - \alpha(0, \widehat{\mathbf{L}}(0))) K(\alpha)^{-1} e \right] \rightarrow \mathbf{E} \left( \int_0^\tau Z(u) du \right) \tag{41}$$

as  $\alpha \rightarrow 0$  through values for which  $K(\alpha)$  is non-singular. Relation (41) amounts to applying L'Hôpital's rule in order to determine the  $(n \times 1)$  vector  $\widehat{\mathbf{h}}_1(0) = \mathbf{E}(\int_0^\tau Z(u) du) = \mathbf{E}(\int_0^\tau I^+(u) du)$ .

For the derivation of  $\widehat{\mathbf{h}}_2(0)$  and similar as above, substituting  $\beta = 0$  in (32) leads to

$$\frac{\partial}{\partial \alpha} \left[ (\mathbb{E}(\mathbf{1}_{\mathcal{J}(T')}) - e^{\alpha q} I) K(\alpha)^{-1} e \right] \rightarrow \mathbb{E} \left( \int_0^{T'} \widetilde{X}(s) ds \right) \tag{42}$$

as  $\alpha \rightarrow 0$  through values for which  $K(\alpha)$  is non-singular. Applying L'Hôpital's rule determines the  $(n \times 1)$  vector  $\widehat{\mathbf{h}}_2(0) = \mathbf{E}(\int_0^{T'} \widetilde{X}(s) ds) = \mathbf{E}(\int_0^{T'} I^-(s) ds)$ .

### 5.2 Long-run average production rate

To compute  $\widehat{\mathbf{p}}_1(0)$  set  $\beta = 0$  in (36). Then, we would like to set  $\alpha = 0$ ; however, this is not viable, as  $\det K(0) = 0$ . Instead we have to use (36) to conclude that

$$(e^{\alpha q} \mathbb{E}(\mathbf{1}_{\mathcal{J}(\tau)}) - I - \alpha(0, \widehat{\mathbf{L}}(0))) K(\alpha)^{-1} \mathbf{C}^+ e \rightarrow \mathbf{E} \left( \int_0^\tau \mathbf{1}_{\mathcal{J}(s)} ds \right) \tag{43}$$

as  $\alpha \rightarrow 0$  through values for which  $\det K(\alpha) \neq 0$ . Relation (43) amounts to applying L'Hôpital's rule in order to determine the vector  $\widehat{\mathbf{p}}_1(0) = \mathbf{E}(\int_0^\tau \mathbf{1}_{\mathcal{J}(s)} ds)$ .

Similarly as above, set  $\beta = 0$  in (38) returns

$$\mathbb{E} \left( \int_0^{T'} e^{\alpha \tilde{X}(s)} \mathbf{1}_{\mathcal{J}(s)} ds \right) = (\mathbb{E} (\mathbf{1}_{\mathcal{J}(T')}) - e^{\alpha q} I) K(\alpha)^{-1}. \tag{44}$$

By setting  $\alpha \rightarrow 0$  and applying L'Hôpital's rule we obtain  $\widehat{\mathbf{u}}_2(0)$ .

### 5.3 Long-run average lost demand

$\widehat{\mathbf{u}}(0)$  has been defined as the  $(n \times 1)$  vector whose  $i$ th component,  $(\widehat{\mathbf{u}}(0))_i$ , represents the expected total demand loss until time  $\tau$  given  $\mathcal{J}_0 = i \in S$ , i.e.,

$$\widehat{\mathbf{u}}(0) = \mathbf{E} \left( \int_0^\tau dL(t) \right) = \mathbf{E} (L(\tau)). \tag{45}$$

**Corollary 5.1** *It is easy to see that  $\widehat{\mathbf{u}}(0) = \widehat{\mathbf{L}}(0)e$ .*

*Remark 3* Notice that  $L_i(\tau)$ , the total demand loss until time  $\tau$  given  $\mathcal{J}_0 = i \in S$ , has a phase type distribution with representation

$$\begin{cases} PH(S_2, e_i \mathbf{0} \widehat{f}_{11}(0, q, 0)e, e_i \mathbf{0} \Psi(0), Q_{22} + Q_{21} \mathbf{0} \Psi(0)) & i \in S_1 \\ PH(S_2, 0, e_i, Q_{22} + Q_{21} \mathbf{0} \Psi(0)) & i \in S_2 \end{cases}.$$

The proof is as follows. We distinguish between two cases: (i)  $i \in S_1$  and (ii)  $i \in S_2$ . In the first case  $e_i \mathbf{0} \widehat{f}_{11}(0, q, 0)e$  is the probability of no loss. According to Table 1,  $e_i \mathbf{0} \Psi(0)$  is the probability vector to hit level 0 avoiding  $q$ . From that point on, there is a demand loss for a time period that has the same distribution as the time to absorption in a CTMC with transition rate matrix  $Q_{22} + Q_{21} \mathbf{0} \Psi(0)$ . (The rate matrix  $Q_{22}$  refers to internal rate changes; the matrix  $Q_{21}$  is the rate matrix to exit level 0 and with probability  $\mathbf{0} \Psi(0)$  the process returns to 0 avoiding  $q$ .) For the second case, the inventory process starts with lost demand; thus the probability of no loss is 0. The rest of the proof is similar.

### 5.4 The expected cycle length

Regarding  $E(C) = E(\tau) + E(T')$ , let  $\mathbf{E}(\tau)$  be the  $(n \times 1)$  vector whose  $i$ th component is  $E_i(\tau) = E(\tau \mid \mathcal{J}_0 = i)$  and, similarly, let  $\mathbf{E}(T')$  be the  $(n \times 1)$  vector whose  $i$ th component is  $E_i(T') = E(T' \mid \mathcal{J}(\tau) = i)$ . Obviously,

$$\begin{aligned} E(\tau) &= \pi^* \mathbf{E}(\tau) \\ E(T') &= \pi^* \mathbb{E} (\mathbf{1}_{\mathcal{J}(\tau)}) \mathbf{E}(T'). \end{aligned} \tag{46}$$

Finally, we will determine  $\mathbf{E}(\tau)$  and  $\mathbf{E}(T')$ .



### 5.4.1 The expected ON period length

In order to obtain  $E_i(\tau)$  (hence, we apply the quantities associated with  $\mathcal{F}^+(t)$ ), we use the notation of (20) and (21) and apply the method introduced in Asmussen (2003), Sect. XI, pp. 312–313. The matrix  $K(\alpha)$  [see (21)] has a real eigenvalue  $k(\alpha)$  with maximal real part (c.f. Asmussen 2003, Chapter XI, p. 312). The corresponding left and right eigenvectors  $v^\alpha$  and  $h^\alpha$  may be chosen with strictly positive components. Moreover, without loss of generality, it can be assumed that  $v^\alpha h^\alpha = 1$ , and that  $\pi h^\alpha = 1$  where  $\pi = v^0$  is the stationary distribution of  $\mathcal{J}(t)$  and  $h^0$  is a column vector of 1's. Let  $h'$  be the derivative of  $h^\alpha$  at  $\alpha = 0$ . Corollary 2.6, Chapter XI, p. 313 of Asmussen (2003) implies that for any stopping time  $\tau$  with  $E_i(\tau) < \infty$ ,

$$E_i(X(\tau)) = k'(0)E_i(\tau) + h'_i - E_i\left(h'_{\mathcal{J}(\tau)}\right). \tag{47}$$

By Corollaries 2.7 and 2.8, Chapter XI, p. 313 of Asmussen (2003):

$$k'(0) = \sum_{i \in S} \pi_i r_i^1. \tag{48}$$

To find the vector  $h'$  we apply the same technique as in Asmussen and Kella (2000, p. 385). We start from

$$k(\alpha)h^\alpha = K(\alpha)h^\alpha. \tag{49}$$

Take the derivative of (49) with respect to  $\alpha$ , let  $\alpha = 0$ . Furthermore, use  $k(0) = 0$ ,  $h^0 = e$ ,  $K'(0) = U$  and  $K(0) = Q$ . This yields

$$k'(0)e = Qh' + Ue. \tag{50}$$

Note that  $\pi h^\alpha = 1$ , so that  $\pi h' = 0$ . Subtracting  $e\pi h' = 0$  from the two sides of (50) shows that

$$h' = (Q - e\pi)^{-1} [k'(0)I - U]e. \tag{51}$$

Now notice that  $Z(t) = X(t) + L(t)$  where  $L(t)$  represents the lost demand until time  $t$ . Thus, applying (47) we obtain

$$E_i(X(\tau)) = E_i(Z(\tau)) - E_i(L(\tau)) = k'(0)E_i(\tau) + h'_i - E_i\left(h'_{\mathcal{J}(\tau)}\right), \tag{52}$$

where

$$\begin{aligned} E_i(Z(\tau)) &= q \\ E_i(h'_{\mathcal{J}(\tau)}) &= e_i \mathbb{E}(\mathbf{1}_{\mathcal{J}(\tau)}) h'. \end{aligned} \tag{53}$$

For the derivation of  $E_i(L(\tau))$  see (45). Substituting (53) in (52), we obtain the  $(n \times 1)$  vector of expectations

$$\mathbf{E}(\tau) = \frac{qe - \widehat{\mathbf{u}}(0) - \mathbf{h}' + \mathbb{E}(\mathbf{1}_{\mathcal{J}(\tau)}) \mathbf{h}'}{k'(0)}. \tag{54}$$

5.4.2 The expected OFF period length

We shift the origin time to  $\tau$  and assume  $\mathcal{J}(\tau) = i \in S$ . Consider  $\tilde{X}(t)_{0 \leq t \leq T'}$  as defined in Sect. 4.2.2 but with  $X(0) = 0$  (clearly with the quantities associated with  $\mathcal{F}^-(t)$ ). It is easy to see that  $\{\tilde{X}(t)\}_{0 \leq t \leq T'}$  has the same distribution as  $\{I(t) - q\}_{0 \leq t \leq T'}$ . Applying again the method of Sect. 5.4.1 to derive  $E_i(T')$  leads to

$$E_i(\tilde{X}(T')) = k'(0)E_i(T') + h'_i - E_i(h'_{\mathcal{J}(T')}), \tag{55}$$

with

$$k'(0) = \sum_{i \in S} \pi_i r_i^2.$$

$$\mathbf{h}' = (Q - e\pi)^{-1} [k'(0)I - U] e.$$

In a similar way we obtain

$$E_i(\tilde{X}(T')) = -q$$

$$E_i(h'_{\mathcal{J}(T')}) = e_i \mathbb{E}(\mathbf{1}_{\mathcal{J}(T')}) \mathbf{h}'. \tag{56}$$

Substituting (56) in (55) leads to the  $(n \times 1)$  vector of expectations

$$\mathbf{E}(T') = \frac{-qe - \mathbf{h}' - \mathbb{E}(\mathbf{1}_{\mathcal{J}(T')}) \mathbf{h}'}{k'(0)}. \tag{57}$$

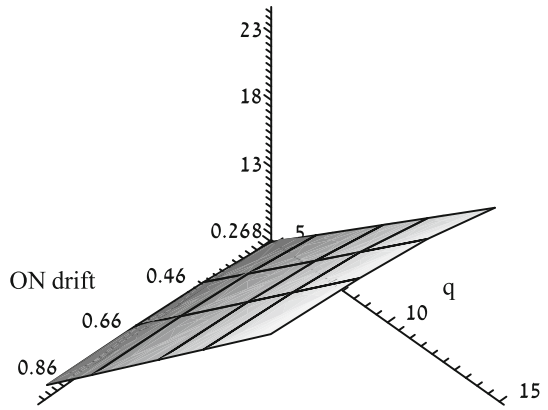
Substituting (57) and (54) in (46) returns the average cycle length.

We summarize the main computation steps of the performance measures for the average case:

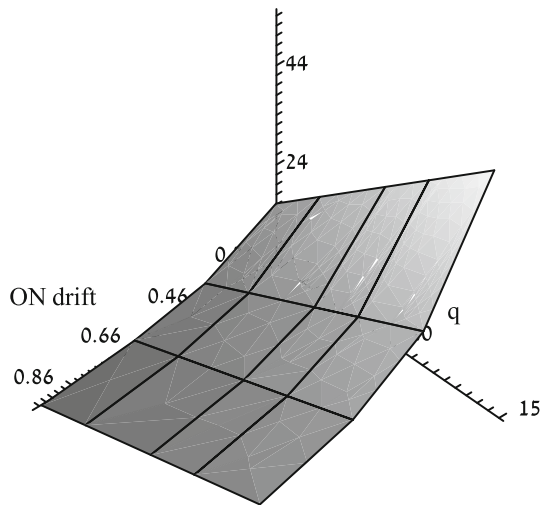
1. Applying (39), (54) and (57) to obtain  $\pi^*$ ,  $\mathbf{E}(\tau)$  and  $\mathbf{E}(T')$ ; then derive  $E(C)$  and  $\overline{SC}$ .
2. Using (41) and (42) to obtain  $\widehat{\mathbf{h}}_1(0)$  and  $\widehat{\mathbf{h}}_2(0)$ ; then compute  $\overline{HC}$ .
3. Similarly, use (43) and (44) to obtain  $\widehat{\mathbf{p}}_1(0)$  and  $\widehat{\mathbf{p}}_2(0)$  which leads to  $\overline{PC}$ .
4. Finally, Corollary 5.1 and (27) leads to  $\widehat{\mathbf{u}}(0)$  and  $\overline{UC}$ .

*Example 2* We consider the same data as in Example 1. This time we vary  $q$  in  $\{5, 7, 10, 12, 15\}$ . Figures 6, 7 and 8 presents the average production cost  $\overline{PC}$ , the average ON period length  $E(\tau)$  and the average OFF period length  $E(T')$ . As expected,  $\overline{PC}$  increases as a function of both  $\gamma_1$  and  $q$ . Figures 6 and 7 shows that both  $E(\tau)$  and  $E(T')$  increases as a function of  $q$ ; however  $E(\tau)$  decreases as a function of  $\gamma_1$  and  $E(T')$  almost unaffected by  $\gamma_1$  since increasing  $\gamma_1$  does not affect on the OFF functionals.

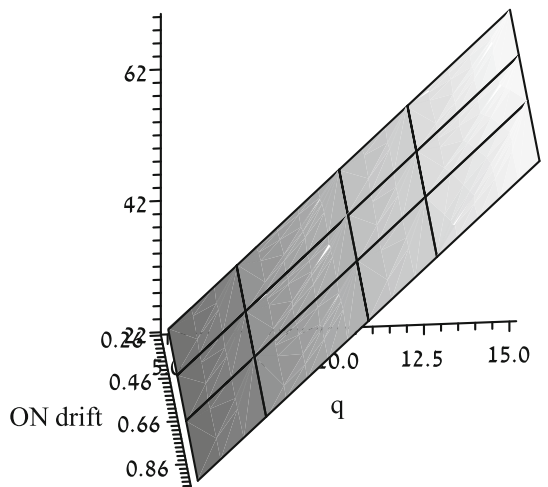
**Fig. 6** The average production cost,  $\overline{PC}$



**Fig. 7** The average ON period length,  $E(\tau)$



**Fig. 8** The average OFF period length,  $E(T')$



**Compliance with ethical standards**

**Conflict of interest** The authors declare that they have no conflict of interest.

**Appendix**

In this ‘‘Appendix’’ we introduce the algorithm developed in Ramaswami (2006) for the computation of the matrix  $\Psi(\beta)$  associated with the MMFF process  $(\mathcal{F}(t), \mathcal{J}(t), t \geq 0)$ , as described in Sect. 3.2. Let  $\sigma(x) = \inf\{t > 0, \mathcal{F}(t) = x\}$  be the first passage time to level  $x$  and define the following LST

$$[\Psi(\beta)]_{ij} = E \left( e^{-\beta\sigma(0)}, \mathcal{J}(\sigma(0)) = j \mid \mathcal{F}(0) = 0, \mathcal{J}(0) = i \right) \quad i \in S_1, \quad j \in S_2. \tag{58}$$

$[\Psi(\beta)]_{ij}$  represents the LST of  $\sigma(0)$  restricted to the event that the fluid process hits level 0 at state  $j \in S_2$ , given  $\mathcal{F}(0) = 0$  and  $J(0) = i \in S_1$ . Ramaswami (2006, Appendix 1, p. 512) shows how to compute the matrix  $\Psi(\beta)$  and provides a good algorithm for this. The algorithm for  $\Psi(\beta)$  is given as follows.

For  $\lambda > 0$ , let

$$P_\lambda = \frac{1}{\lambda}U^{-1}Q + I$$

where  $U$  and  $Q$  define in Sect. 3.2. Choose (fixed) positive numbers  $\lambda$  and  $\delta$  such that

$$\begin{aligned} \lambda &\geq \max_{i \in S} \left\{ -[U^{-1}Q]_{ii} \right\} \\ \max_{i \in S} \left[ \frac{\text{Re}(\beta)}{\lambda}U^{-1} \right]_{ii} &\leq \delta < 1, \quad \text{and} \\ \max_{i \in S} \left[ P_\lambda - \frac{\text{Re}(\beta)}{\lambda}U^{-1} \right]_{ii} &> 0 \end{aligned}$$

Define the matrices

$$\begin{aligned} A_2(\beta, \lambda) &= \begin{pmatrix} 0 & 0 \\ 0 & \lambda U_2 (\beta I + 2\lambda U_2)^{-1} \end{pmatrix} \\ A_1(\beta, \lambda) &= \Lambda U (\beta I + \Lambda U)^{-1} \begin{pmatrix} 0 & 0 \\ \frac{1}{2}P_{21} & \frac{1}{2}P_{22} \end{pmatrix} \\ A_0(\beta, \lambda) &= \begin{pmatrix} P_{11} - \frac{\beta}{\lambda}U_1^{-1} & P_{12} \\ 0 & 0 \end{pmatrix} \end{aligned}$$

where  $\Lambda = \text{diag}(\lambda I_{|S_1|}, 2\lambda I_{|S_2|})$  and  $P = P_\lambda$ . Consider now the following algorithm.

**Algorithm**

Fix  $\epsilon > 0$  and set  $diff = 100$ ;

$$\begin{aligned}
 H^{**}(1, \beta, \lambda) &= (I - A_1(\beta, \lambda))^{-1} A_0(\beta, \lambda); \\
 L^{**}(1, \beta, \lambda) &= (I - A_1(\beta, \lambda))^{-1} A_2(\beta, \lambda); \\
 G^{**}(1, \beta, \lambda) &= L^{**}(1, \beta, \lambda); \\
 T(1) &= H^{**}(1, \beta, \lambda); \\
 \text{Dowhile}(diff > \epsilon) \\
 &k = k + 1; \\
 U^{**}(k, \beta, \lambda) &= H^{**}(k - 1, \beta, \lambda)L^{**}(k - 1, \beta, \lambda) \\
 &\quad + L^{**}(k - 1, \beta, \lambda)H^{**}(k - 1, \beta, \lambda); \\
 M &= (H^{**}(k - 1, \beta, \lambda))^2; \\
 H^{**}(k, \beta, \lambda) &= (I - U^{**}(k, \beta, \lambda))^{-1}M; \\
 M &= (L^{**}(k - 1, \beta, \lambda))^2; \\
 L^{**}(k, \beta, \lambda) &= (I - U^{**}(k, \beta, \lambda))^{-1}M; \\
 G^{**}(k, \beta, \lambda) &= G^{**}(k - 1, \beta, \lambda) + T(k - 1)L^{**}(k, \beta, \lambda); \\
 T(k) &= T(k - 1)H^{**}(k, \beta, \lambda); \\
 diff &= \max_{j,k \in S} \{ [G^{**}(k, \beta, \lambda)]_{j,k} - [G^{**}(k - 1, \beta, \lambda)]_{j,k} \}; \\
 \text{end} \\
 \Psi(\beta) &\cong G_{12}^{**}(k, \beta, \lambda)G_{22}^{**}(k, \beta, \lambda)^{-1}.
 \end{aligned}$$

Once we have computed  $\Psi(\beta)$ , the LST matrix of other hitting times are straightforward to evaluate; we list these matrices and their sizes in Tables 3 and 4. All matrices have nice probabilistic interpretations. For more details see Ahn et al. (2007).

The following LSTs of first-passage times are needed in our analysis:

**Table 3** Transform matrices

Quantity	Matrix size
$K(\beta) = U_1^{-1}(Q_{11} - \beta I) + \Psi(\beta)U_2^{-1}Q_{21}$	$n_1 \times n_1$
$H(\beta) = U_2^{-1}(Q_{22} + Q_{21}\Psi(\beta))$	$n_2 \times n_2$
$\Xi(\beta, x) = \Psi(\beta) \int_{(0,x)} e^{H(\beta)y} (U_2^{-1}Q_{21}^1) e^{K(\beta)y} dy$	$n_1 \times n_1$

**Table 4** LST of first-passage times

LST	First passage time	Matrix size
${}_0\hat{f}_{12}(x, 0, \beta) = \Psi(\beta)e^{H(\beta)x}$	From $(x, S_1)$ to $(0, S_2)$ avoiding 0 in $\mathcal{F}$	$n_1 \times n_2$
${}_0\hat{f}_{22}(x, 0, \beta) = e^{H(\beta)x}$	From $(x, S_2)$ to $(0, S_2)$ avoiding 0 in $\mathcal{F}$	$n_2 \times n_2$
${}_0\hat{f}_{11}(0, x, \beta) = e^{K(\beta)x} (I + \Xi(\beta, x))^{-1}$	From $(0, S_1)$ to $(x, S_1)$ avoiding 0 for $\mathcal{F}$	$n_1 \times n_1$
${}_x\Psi(\beta) = \Psi(\beta) - {}_0\hat{f}_{11}(0, x, \beta)\Psi(\beta)e^{H(\beta)x}$	From $(0, S_1)$ to $(0, S_2)$ avoiding $x$ for $\mathcal{F}$	$n_1 \times n_2$

## References

- Ahn S, Badescu AL, Ramaswami V (2007) Time dependent analysis of finite buffer fluid flows and risk models with a dividend barrier. *Queueing Syst* 55:207–222
- Asmussen S (2003) *Applied probability and queues*, 2nd edn. Springer, New York
- Asmussen S, Kella O (2000) A multi-dimensional martingale for Markov additive process and its applications. *Adv Appl Probab* 32:376–393
- Barron Y (2015) A fluid EOQ model with Markovian environment. *J Appl Probab* 52:473–489
- Barron Y, Perry D, Stadje W (2014) A make-to-stock production/inventory model with MAP arrivals and phase-type demands. *Ann Oper Res* 1–37
- Beltran JL, Krass D (2002) Dynamic lot sizing with returning items and disposals. *IIE Trans* 34:437–448
- Berman O, Perry D (2001) Two control policies for stochastic EOQ-type models. *Probab Eng Inf Sci* 15:445–463
- Berman O, Perry D, Stadje W (2007) Performance analysis of a fluid production/inventory model with state-dependence. *Methodol Comput Appl Probab* 9:465–481
- Boxma O, Kaspi H, Kella O, Perry D (2005) On/off storage systems with state-dependent input, output, and switching rates. *Probab Eng Inf Sci* 19:1–14
- Boxma OJ, Kella O, Perry D (2001) An intermittent fluid system with exponential on-times and semi-Markov input rates. *Probab Eng Inf Sci* 15:189–198
- Boxma O, Perry D, Stadje W, Zacks S (2015) A compound Poisson EOQ model for perishable items with intermittent high and low demand periods. Working paper (in press)
- Boxma O, Perry D, Zacks S (2014) A fluid EOQ model of perishable items with intermittent high and low demand rates. *Math Oper Res* 40:390–402
- Da Silva Soares A, Latouche G (2005) Matrix-analytic methods for fluid queues with feedback control. *Int J Simul Syst Sci Technol* 6:4–12
- De Kok AG, Tijms HC, Van Der Duyn Schouten FA (1984) Approximations for the single product production–inventory problem with compound Poisson demand and service level constraints. *Adv Appl Probab* 16:378–401
- Fleischmann M, Kuik R, Dekker R (2002) Controlling inventories with stochastic item returns: a basic model. *Eur J Oper Res* 138:63–75
- Germers R, Forest NDV (2014) Optimal control of production–inventory systems with constant and compound Poisson demand. University of Groningen, Groningen
- Grunow M, Günter HO, Westinner R (2007) Supply optimization for the production of raw sugar. *Int J Prod Econ* 110:224–239
- Kella O, Perry D, Stadje W (2003) A stochastic clearing model with Brownian and a compound Poisson component. *Probab Eng Inf Sci* 17:1–22
- Kella O, Whitt W (1992) A Storage model with a two-stage random environment. *Oper Res* 40:s257–s262
- Kulkarni VG, Yan K (2007) A fluid model with upward jumps at the boundary. *Queueing Syst* 56:103–117
- Latouche G, Ramaswami V (1999) Introduction to matrix analytic methods in stochastic modeling, vol 5. Society for Industrial and Applied Mathematics (SIAM), Philadelphia
- Mohebbi E (2006) A production–inventory model with randomly changing environmental conditions. *Eur J Oper Res* 174:539–552
- Perry D, Posner MJ (2002) A mountain process with state dependent input and output and a correlated dam. *Oper Res Lett* 30:245–251
- Pinçe Ç, Gürler Ü, Berk E (2008) A continuous review replenishment–disposal policy for an inventory system with autonomous supply and fixed disposal costs. *Eur J Oper Res* 190:421–442
- Ramaswami V (1999) Matrix analytic methods for stochastic fluid flows. In: *Proceedings of the international teletraffic congress (ITC-16)*. Edinburgh, Elsevier
- Ramaswami V (2006) Passage times in fluid models with application to risk processes. *Methodol Comput Appl Probab* 8:497–515
- Shaharudin MR, Zailani S, Tan KC (2015) Barriers to product returns and recovery management in a developing country: investigation using multiple methods. *J Clean Prod* 96:220–232
- Shen Y, Willems SP (2014) Modeling sourcing strategies to mitigate part obsolescence. *Eur J Oper Res* 246:522–533
- Shi J, Katehakis MN, Melamed B (2013) Martingale methods for pricing inventory penalties under continuous replenishment and compound renewal demands. *Ann Oper Res* 208:593–612

- Shi J, Katehakis MN, Melamed B, Xia Y (2014) Production-inventory systems with lost-sales and compound Poisson demands. *Oper Res* 62:1048–1063
- Song Y, Lau HC (2004) A periodic-review inventory model with application to the continuous-review obsolescence problem. *Eur J Oper Res* 159:110–120
- Song JS, Zipkin P (1993) Inventory control in a fluctuating demand environment. *Oper Res* 41:351–370
- Taft EA (1918) The most economical production lot. *Iron Age* 101(18):1410–1412
- Taylor HM (1968) Some models in epidemic control. *Math Biosci* 3:383–398
- Tyler R (2004) Industry confronts equipment crunch. *The Daily Telegraph* (London) May 6
- Vickson RG (1986) A single product cycling problem under Brownian motion demand. *Manag Sci* 32:1336–1345
- Wu J, Chao X (2013) Optimal control of a Brownian production/inventory system with average cost criterion. *Math Oper Res* 39:163–189
- Zhang B, Zwart B (2012) Fluid models for many-server Markovian queues in a changing environment. *Oper Res Lett* 40:573–577