

# A JUMP-FLUID PRODUCTION–INVENTORY MODEL WITH A DOUBLE BAND CONTROL

YONIT BARRON

*Department of Statistics  
University of Haifa  
Haifa 91905  
Israel  
Email: [ybarron@stat.haifa.ac.il](mailto:ybarron@stat.haifa.ac.il)*

DAVID PERRY

*Department of Statistics  
University of Haifa  
Haifa 91905  
Israel  
Email: [dperry@stat.haifa.ac.il](mailto:dperry@stat.haifa.ac.il)*

WOLFGANG STADJE

*Institute of Mathematics  
University of Osnabrück  
49069 Osnabrück  
Germany  
Email: [wstadje@uos.de](mailto:wstadje@uos.de)*

We consider a production–inventory control model with two reflecting boundaries, representing the finite storage capacity and the finite maximum backlog. Demands arrive at the inventory according to a Poisson process, their i.i.d. sizes having a common phase-type distribution. The inventory is filled by a production process, which alternates between two prespecified production rates  $\rho_1$  and  $\rho_2$ : as long as the content level is positive,  $\rho_1$  is applied while the production follows  $\rho_2$  during time intervals of backlog (i.e., negative content). We derive in closed form the various cost functionals of this model for the discounted case as well as under the long-run-average criterion. The analysis is based on a martingale of the Kella–Whitt type and results for fluid flow models due to Ahn and Ramaswami.

## 1. INTRODUCTION

An important problem in production planning is to select the production rates, the storage capacity and the backlog possibilities in order to cope with random fluctuation in the demands. Determining how fast the production should respond to the randomly incoming demands and to what extent their variability should be absorbed by accumulating

inventory depends on the relative importance of various associated costs. One has to take into account the cost rates for production, holding of inventory, backlogging, lost sales and lost productivity. Fixing the production rate at one constant high level avoids backlogs but causes high production and inventory costs. On the other hand, fixing the production rate at a constant low level may lead to severe backlogging and lost sale opportunities.

In this paper, we consider a production–inventory model that combines a standard compound Poisson demand arrival process with the following special features:

- (a) the storage capacity is finite;
- (b) backlog is permitted up to a threshold;
- (c) there are two different production rates for the periods of positive and of negative inventory content levels.

Specifically, the production rate is low (say  $\rho_1$ ) when the inventory level is above zero (so that the inventory is not filled too quickly) and high (say  $\rho_2$  with  $\rho_2 > \rho_1$ ) when the inventory level is below zero (to reduce backorders and lost sales). We assume that a switch from one production rate to the other takes no time. The demands arrive according to a Poisson process with rate  $\lambda$ , and their sizes are independent random variables with an arbitrary common phase-type distribution. If the storage facility is filled to capacity, the production is stopped until a new demand arrives. Any demand which cannot be satisfied immediately is backlogged, provided the total backlog does not exceed a given level; otherwise the excess demand is lost.

In order to manage such a production–inventory model, one needs information about the holding cost for the stock, the cost of lost production due to the finite storage capacity, the shortage cost for the backlogged demand and the cost due to unsatisfied demand. The main objective of this paper is to develop techniques to determine explicit formulas for all these costs under the discounted as well as under the long-run average cost criterion.

For background on production–inventory control models see Doshi, Schouten and Talman [8], in particular with regard to the long-run average cost criterion. Perry et al. [13] studied the behavior of a broker in a dealership market whose buffer content is governed by stochastically dependent demand and supply. Another related model is the so-called clearing system (see e.g. Boxma et al. [6], Kella et al. [10] and Perry et al. [14]), which can be regarded as the dual stochastic Economic Order Quantity (EOQ) model. In a clearing system, the content process jumps back to zero when it reaches a certain positive level.

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 Kella and Whitt [9] and was frequently used in the study of inventory models, see e.g. Perry et al. [13] and Kella et al. [10] and references given therein.

Fluid flows 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. A standard example of a fluid flow is given by an infinite capacity buffer with inflow and outflow rates controlled by a Markov chain. The buffer level increases or decreases linearly at the current rate. When it becomes empty, several strategies can be applied: it can remain empty until the demand level reaches a certain barrier (see e.g. Baek et al. [5]) or it can have positive jumps at the boundary. Kulkarni and Yan [11,12] and Yan and Kulkarni [16] studied inventory models with instant stock replenishments. They derive a system of first-order non-homogeneous linear differential equations for the limiting distribution and an EOQ policy that minimizes the long-run average cost. They also expand the model to allow backlogging and exponential leadtimes. Ramaswami [15] and Ahn and Ramaswami [4,15] presented a unified methodology for studying a large class of insurance risk models via fluid

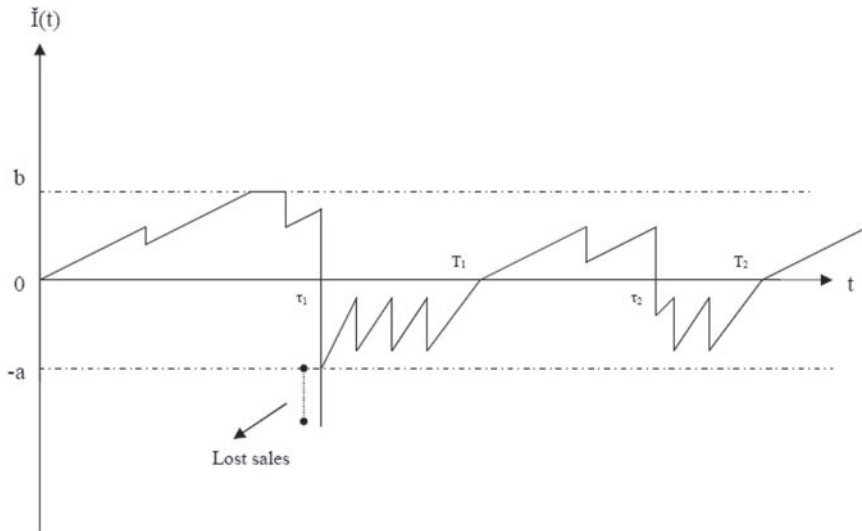


FIGURE 1. A typical sample path of the inventory level process.

flows by making use of the connection between an insurer’s surplus process and a particular fluid flow. To the best of our knowledge, fluid flow methods have not yet been applied to production–inventory models and the associated cost-related quantities.

The paper is organized as follows. In Section 2, we present the mathematical description of the model and the cost functionals. The crucial tools of our analysis are introduced in Section 3. In Section 4, we derive all the cost functionals in closed form. A numerical example is given in Section 5. Finally, the long-run average case is considered in Section 6.

**2. MATHEMATICAL DESCRIPTION OF THE MODEL**

The counting process of the demand arrival times is a Poisson process  $\{N(t), t \geq 0\}$  with rate  $\lambda$ . We assume that the demand amounts  $V_1, V_2, \dots$ , are i.i.d. and have a phase-type distribution with representation  $(\epsilon, \pi, G)$ . This is the distribution of the time until absorption for a continuous-time Markov chain with a finite set  $\epsilon$  of transient states and one absorbing state.  $\pi$  is the vector of the initial probabilities for the transient states and  $G$  is the transition rate matrix among the transient states. We set  $\epsilon = \{1, \dots, n\}$ . Note that  $G^o = -Ge$  is the vector of transition rates to the absorbing state (where  $e$  is the vector with all its  $n$  components equal to 1) and that the Laplace transform and the mean of  $V = V_1$  are given by  $M_V(\alpha) = \pi(\alpha I - G)^{-1}G^o$  and  $E(V) = -\pi G^{-1}e$ .

The storage capacity  $b$  is assumed to be finite. If the inventory is filled to capacity the production is stopped, until a new demand arrives. We also suppose that backordering up to a certain level  $-a$  is allowed ( $a > 0$ ). Any backlog that exceeds this level is lost, and the replenishment of the inventory continues from level  $-a$ .

The inventory is filled, by production, at rate  $\rho_1$ , while its level is nonnegative and at rate  $\rho_2$ , while its level is negative. Since the demand sizes have a continuous distribution, the only way to get to level 0 is from below (with probability 1), and at that time the rate changes from  $\rho_2$  to  $\rho_1$ .

Let  $\check{I}(t)$  be the (double reflected) *inventory level process*. A typical sample path of  $\check{I}(t)$  is given in Figure 1. The process satisfies the equation

$$\check{I}(t) = \max \left[ -a, \check{I}(0) + \int_0^t \rho_1 \mathbf{1}_{\{\check{I}(s) \in [0, b]\}} ds + \int_0^t \rho_2 \mathbf{1}_{\{\check{I}(s) < 0\}} ds - \sum_{i=1}^{N(t)} V_i \right], \quad t \geq 0, \quad (1)$$

where  $\mathbf{1}_A$  is the indicator of the event  $A$ .

Clearly, if  $a < \infty$  and  $b < \infty$  the inventory level process is stable (its expected return time to level 0 is finite). In the case  $a = \infty$  and  $b = \infty$ , the inequalities

$$\rho_1 < \lambda E(V) < \rho_2 \quad (2)$$

are necessary and sufficient for stability. If only one of  $a$  and  $b$  is infinite, stability is equivalent to the validity of the corresponding one of the two inequalities. Thus, if (2) holds,  $\check{I}(t)$  is stable for all values of  $a$  and  $b$ . The first inequality ensures that high holding cost is avoided, while the second one is needed to avoid high shortage cost.

The process  $\check{I}(t)$  can be partitioned into the *positive process*  $\check{I}^+(t) = \max[\check{I}(t), 0]$  and the *negative process*  $\check{I}^-(t) = -\min[0, \check{I}(t)] = \check{I}^+(t) - \check{I}(t)$ . During their respective continuous segments,  $\check{I}^+(t)$  and  $-\check{I}^-(t)$  increase at rates  $\rho_1$  and  $\rho_2$ , respectively.

Owing to the drift condition (2), after the inventory process becomes negative, it will upcross level zero with probability 1 and then continue at rate  $\rho_1$ . The time instants of these rate switchings are called *recovery points*:

$$T_0 = 0 \quad \text{and} \quad T_n = \inf\{t > T_{n-1} : \check{I}(t) = 0\}, \quad n \geq 1.$$

They form a renewal process, and the inventory level process is regenerative with i.i.d. cycles during  $[T_{n-1}, T_n)$ . In every cycle  $\check{I}(t)$  switches from rate  $\rho_1$  to rate  $\rho_2$  exactly once; between  $T_{n-1}$  and  $T_n$  this switching occurs at time

$$\tau_n = \inf\{t > T_{n-1} : \check{I}(t) < 0\}. \quad (3)$$

We call the times  $\tau_n$  the *zero points*. We will mainly consider the first cycle and write  $\tau$  and  $T$  for  $\tau_1$  and  $T_1$ .

Let us now introduce the functionals indicating the *expected discounted costs* in our model, using the discount factor  $\beta > 0$ .

(a) *Holding cost.* The expected total discounted holding cost can be expressed as

$$HC(\beta) = hE \int_0^\infty e^{-\beta t} \check{I}^+(t) dt, \quad (4)$$

where  $h dt$  is the holding cost for a unit of stock during an infinitesimal time interval of length  $dt$ . Let  $\hat{h}(\beta) = E \left( \int_0^\tau e^{-\beta t} \check{I}(t) dt \right)$ . Invoking the ergodic theorem for regenerative processes, we can write  $HC(\beta)$  in terms of the first cycle as

$$HC(\beta) = \frac{h\hat{h}(\beta)}{1 - E(e^{-\beta T})}.$$

(b) *Idle time cost.* Owing to the finite storage capacity  $b$ , the production is stopped at this level. As a measure for the expected discounted loss due to idle times of the production

facility one can use the functional

$$KC(\beta) = kE \int_0^\infty \rho_1 e^{-\beta t} \mathbf{1}_{\{\check{I}^+(t)=b\}} dt, \tag{5}$$

where  $k dt$  is the cost of production loss, due to limited capacity, during a time interval of length  $dt$ . Let  $\widehat{k}(\beta) = E \left( \int_0^\tau e^{-\beta t} \mathbf{1}_{\{\check{I}(t)=b\}} dt \right)$ . In terms of the first cycle, we have

$$KC(\beta) = \frac{k\rho_1 \widehat{k}(\beta)}{1 - E(e^{-\beta T})}. \tag{6}$$

(c) *Shortage cost.* The expected total discounted shortage cost is

$$SC(\beta) = -wE \int_0^\infty e^{-\beta t} \check{I}^-(t) dt, \tag{7}$$

where  $w dt$  is the penalty cost for a unit of shortage during a time interval of length  $dt$ . We use  $\mathbf{E}(e^{-\beta\tau})$  as shorthand notation for the  $1 \times n$  vector whose  $i$  component is given by

$$E_i(e^{-\beta\tau}) = E(e^{-\beta\tau} \mathbf{1}_{\{\text{level 0 hit at time } \tau \text{ in phase } i\}}),$$

which is the Laplace transform of the time until *zero point* restricted to the event that the process hits level 0 with phase  $i \in \epsilon$ . Moreover, let  $\widehat{\mathbf{s}}(\beta)$  be the  $n \times 1$  vector whose  $i$ th component is

$$\widehat{\mathbf{s}}_i(\beta) = \mathbf{E}_i \left( - \int_0^{T-\tau} e^{-\beta t} \check{I}(t + \tau) dt \right),$$

i.e., the expected discounted shortage level, given the initial phase at the zero point is  $i \in \epsilon$ . Then we have, in terms of scalar products,

$$SC(\beta) = w \frac{\mathbf{E}(e^{-\beta\tau}) \mathbf{E} \left[ - \int_0^{T-\tau} e^{-\beta t} \check{I}(t + \tau) dt \right]}{1 - E(e^{-\beta T})} = \frac{w \mathbf{E}(e^{-\beta\tau}) \widehat{\mathbf{s}}(\beta)}{1 - E(e^{-\beta T})}. \tag{8}$$

(d) *Cost of unsatisfied demands.* In our model, the backlog that exceeds level  $(-a)$  is lost. Let  $\varphi$  be the penalty cost per unsatisfied unit of demand. An unsatisfied demand occurs whenever the process  $\check{I}^-(t)$  is down-crossing level  $(-a)$ . Denote the  $i$ th time that this is happening by  $\xi_i$  and the corresponding unsatisfied demand size by  $Y_i$  (i.e.,  $Y_i = \check{I}^-(\xi_i) - (-a)$ , see Figure 1). The expected total discounted cost of the unsatisfied demand can be expressed as

$$UC(\beta) = \varphi E \sum_i e^{-\beta \xi_i} |Y_i|. \tag{9}$$

Let  $N_Y(T)$  be the number of times a demand is unsatisfied in the first cycle and

$$\widehat{u}(\beta) = E \left[ \sum_{i=1}^{N_Y(T)} e^{-\beta \xi_i} |Y_i| \right].$$

Regenerative theory yields

$$UC(\beta) = \frac{\varphi \widehat{u}(\beta)}{1 - E(e^{-\beta T})}. \tag{10}$$

Note that the functionals in (a) and (b) can be defined in terms of the positive process  $(\check{I}^+(t))$ , while the two in (c) and (d) depend only on the negative process  $(\check{I}^-(t))$ .

A simple cost function for the system would be the sum,  $TC(\beta)$ , of these four expected discounted costs, i.e.,

$$TC(\beta) = HC(\beta) + KC(\beta) + SC(\beta) + UC(\beta).$$

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

$$\overline{TC} = \lim_{\beta \rightarrow \infty} \beta TC(\beta) = hE(H_c) + kE(K_c) + wE(S_c) + \varphi E(U_c). \tag{11}$$

In (11),  $E(H_c)$ ,  $E(K_c)$ ,  $E(S_c)$  and  $E(U_c)$  denote the expected values, in steady state, of the inventory level, the production loss due to idleness, the back-ordering level and the amount of unsatisfied demands, respectively. The four long-run average costs are given in terms of first-cycle functionals as follows:

$$E(H_c) = \lim_{\beta \rightarrow 0} \frac{E\left(\int_0^\tau e^{-\beta t} \check{I}(t) dt\right)}{E(T)} = \frac{E\left(\int_0^\tau \check{I}(t) dt\right)}{E(T)} = \frac{\hat{h}(0)}{E(T)}, \tag{12}$$

$$E(K_c) = \lim_{\beta \rightarrow 0} \frac{\rho_1 E\left(\int_0^\tau e^{-\beta t} \mathbf{1}_{\{\check{I}(t)=b\}} dt\right)}{E(T)} = \frac{\rho_1 E\left(\int_0^\tau \mathbf{1}_{\{\check{I}(t)=b\}} dt\right)}{E(T)} = \frac{\rho_1 \hat{k}(0)}{E(T)}, \tag{13}$$

$$E(S_c) = \lim_{\beta \rightarrow 0} \frac{-E\left(\int_\tau^T e^{-\beta t} \check{I}(t) dt\right)}{E(T)} = \frac{-E\left(\int_\tau^T \check{I}(t) dt\right)}{E(T)} = \frac{\hat{s}(0)}{E(T)}, \tag{14}$$

$$E(U_c) = \lim_{\beta \rightarrow 0} \frac{E\left(\sum_{i=1}^{N_Y(T)} e^{-\beta \xi_i} |Y_i|\right)}{E(T)} = \frac{E\left(\sum_{i=1}^{N_Y(T)} |Y_i|\right)}{E(T)} = \frac{\hat{u}(0)}{E(T)}. \tag{15}$$

The explicit formulas derived below for all these quantities also cover the case of the unreflected (or partially reflected) level content process, i.e.,  $b = \infty$  or  $a = \infty$ , in which the inventory level is unbounded or no demand portions are lost.

### 3. MATHEMATICAL PRELIMINARIES

For the determination of the cost functionals we use two tools: (a) the matrix-analytic approach and the theory of Markov-modulated fluid flows (initiated in a series of papers by Ahn and Ramaswami [1]– [3] and (b) an application of the optional sampling theorem to a special Kella–Whitt martingale.

#### 3.1. The Fluid Inventory Model

Consider first the inventory level process above with no backlog and with infinite capacity ( $a = 0, b = \infty$ ) and with only one rate  $\rho$ , i.e.,  $\rho_1 = \rho_2$ . (note that, in this case, the process is either transient or null recurrent). We transform this process pathwise into a fluid flow by replacing each (instantaneous) demand of size  $x$  by a segment of length  $x/\rho$  of linear decrease at rate  $\rho$  per unit time. We call  $\mathcal{F}(t)$  the fluid inventory process generated this way. It can be described as follows. Introduce a modulating continuous-time Markov chain  $\mathcal{J}(t)$  with state space  $S = S_1 \cup S_2$ , where  $S_1 = \{0\}$  and  $S_2 = \epsilon = \{1, 2, \dots, n\}$ . Its infinitesimal

generator  $Q$  is given in a block form according to transitions between the sets  $S_i$  ( $i = 1, 2$ )

$$Q = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix}.$$

Now let  $\mathcal{F}(t)$  be the level of the fluid at time  $t$  that is modulated as follows: whenever the Markov chain is in  $S_1 = \{0\}$  the fluid flow increases at rate  $\rho$ , and whenever it is in  $S_2$ ,  $\mathcal{F}(t)$  decreases at rate  $\rho$ .

Three variants of the fluid flow  $\mathcal{F}$  are particularly useful in the analysis of our inventory level process:

*Finite buffer fluid flow:* We define the finite buffer fluid flow  ${}^b\mathcal{F}$ , which allows the fluid level to decrease only when it is positive and to increase only when it is less than the buffer level  $b > 0$ .

*Reflected fluid flow:* The reflected fluid flow  $\mathcal{F}^r$  is obtained by reversing the roles of the up and down environment states.

*Reflected finite buffer fluid flow:* In a similar manner, we define the reflected finite buffer fluid flow  ${}^b\mathcal{F}^r$ .

In order to apply the fluid inventory model to our original process  $\check{I}(t)$ , we consider the two fluid models  $\mathcal{F}_1$  and  $\mathcal{F}_2$  corresponding to the positive and the negative inventory process, respectively. For  $i = 1, 2$  the  $(1 + n) \times (1 + n)$  generator matrix  $Q_i$  is given by

$$Q_i = \begin{pmatrix} -\lambda & \lambda\pi \\ \rho_i G^o & \rho_i G \end{pmatrix}.$$

All the matrices and the quantities below have to be evaluated for  $\mathcal{F}_1$  and for  $\mathcal{F}_2$ . The computation in both cases is similar, by changing  $\rho$  to  $\rho_1$  or  $\rho_2$ , and  $b$  to  $a$ , respectively.

In order to apply results for fluid models to our original process  $\check{I}(t)$ , we make appropriate changes in the clock time. Let  $\sigma(0, 0)$  be the first passage time from level 0 to level 0 in the fluid inventory model. The following observations are follow from the fact that that for the fluid model, upward and downward rates are all equal in absolute value to a common constant  $\rho$  :

- (a) The first passage time from level 0 to level 0 in the jump inventory model, say  $v(0, 0)$  is just half of the corresponding time in the fluid inventory model  $\mathcal{F}$ , so that we have

$$v(0, 0) = \sigma(0, 0)/2.$$

- (b) For the corresponding intervals of descent from  $(x, S_1)$  to  $(y, S_2)$  with  $0 \leq y < x$ , we have

$$\sigma(x, y) = 2v(x, y) + \frac{x - y}{\rho}.$$

Note that  $(x - y)/\rho$  is the time in the fluid process to descend form  $x$  to  $y$ . Thus, we get  $v(x, y) = \frac{\sigma(x, y)}{2} - \frac{x - y}{2\rho}$ .

- (c) For the corresponding intervals of ascent from  $(x, S_1)$  to  $(y, S_1)$  with  $0 \leq x < y$ , we have

$$v(x, y) = \frac{\sigma(x, y)}{2} + \frac{y - x}{2\rho}.$$

This again in a consequence of the fact that all rates for the fluid model are equal in absolute value.

TABLE 1. Transform matrices.

Quantity	Matrix size
$K(s) = \rho^{-1}(Q_{11} - sI) + \Psi(s)\Theta^{-1}Q_{21}$	$1 \times 1$
$H(s) = \Theta^{-1}(Q_{22} - sI + Q_{21}\Psi(s))$	$n \times n$
$\Xi(s, x) = \Psi(s) \int_{(0,x)} e^{H(s)y} (\Theta^{-1}Q_{21}) e^{K(s)y} dy$	$1 \times 1$

TABLE 2. Transform matrices for the rate-reversed process.

Quantity	Matrix size
$K^r(s) = \Theta^{-1}(Q_{22} - sI) + \Psi^r(s)\rho^{-1}Q_{12}$	$n \times n$
$H^r(s) = \rho^{-1}(Q_{11} - sI + Q_{12}\Psi^r(s))$	$1 \times 1$
$\Xi^r(s, x) = \Psi^r(s) \int_{(0,x)} e^{H^r(s)y} (\rho^{-1}Q_{12}) e^{K^r(s)y} dy$	$n \times n$

Let  $\Psi(s)$  be the  $(1 \times n)$  vector whose  $j$ th component is

$$E(e^{-s\sigma(0,0)} \mathbf{1}_{\{\mathcal{J}(\tau)=j\}} \mid \mathcal{F}(0) = 0),$$

which is the Laplace–Stieltjes transform (LST) of  $\sigma(0,0)$  restricted to the event that the demand hits level 0 at phase  $j$  and given that  $\mathcal{F}(0) = 0$ . Ramaswami ([15], Appendix 1) shows how to compute  $\Psi(s)$  and provides a good algorithm for this. Analogously,  $\Psi^r(s)$  is the vector (of order  $n \times 1$ ) whose  $j$ th component is the LST of the time to reach level 0 for the process  $\mathcal{F}^r$ , given that  $\mathcal{F}^r(0) = 0$ ,  $\mathcal{J}^r(0) = j$ , where  $\mathcal{J}^r(t)$  is the modulated state process for  $\mathcal{F}^r$ .

To apply  $\Psi(s)$  for the original process, we use  $\Psi(\frac{s}{2})$ . Notice that  $\Psi(0)$  is a  $1 \times n$  probability vector. We write  $\Psi_1$  or  $\Psi_2$  if we refer to  $\mathcal{F}_1$  or  $\mathcal{F}_2$ , respectively. Owing to the different drifts, it is obvious that  $\Psi_1(0)\mathbf{1} = 1$  and  $\Psi_2^r(0) = e$ .

Let  $\Theta = \text{diag}(\rho)$  be the  $(n \times n)$  matrix with diagonal element  $\rho$ . All other LST matrices for the hitting times that we will use are straightforward to evaluate once we have computed  $\Psi(s)$ . We list these matrices and their dimensions in Tables 1–3. All matrices have nice probabilistic interpretations. For more details see Ramaswami [15] and Ahn and Ramaswami [4].

Analogously to the matrices considered above for  $\mathcal{F}$ , we also introduce the matrices  $\Psi^r(s)$ ,  $K^r(s)$ ,  $H^r(s)$ ,  $\Xi^r(s)$  associated to the rate-reversed flow  $\mathcal{F}^r$ , by changing the index from 1 to 2 and from 2 to 1.

### 3.2. The Basic Martingale

Let  $X(t)$  be a Lévy process with no negative jumps and Lévy exponent  $\varphi(\alpha) = \log Ee^{-\alpha X(1)}$ . Let  $\{Y(t), t \geq 0\}$  be an adapted process with finite expected variation on finite intervals, and let  $Z(t) = X(t) + Y(t)$ . Kella and Whitt ([9], Theorem 2) have shown that the process

$$M(t, \alpha) = \varphi(\alpha) \int_0^t e^{-\alpha Z(s)} ds + e^{-\alpha Y(0)} - e^{-\alpha Z(t)} - \alpha \int_0^t e^{-\alpha Z(s)} dY(s), \quad t \geq 0 \quad (16)$$

is a martingale. Some of the relevant functionals in this paper will be obtained by applying the martingale stopping theorem to appropriate special cases of (16). For our model we



TABLE 3. LSTs of the first passage times.

<i>LST</i>	<i>First passage time</i>	<i>Matrix size</i>
${}_0\widehat{f}_{22}^r(0, x, s) = e^{K^r(s)x}(I + \Xi^r(s, x))^{-1}$	from $(0, S_2)$ to $(x, S_2)$ avoiding 0 in $F^r$	$n \times n$
${}_0\widehat{f}_{11}(0, x, s) = e^{K(s)x}(I + \Xi(s, x))^{-1}$	from $(0, S_1)$ to $(x, S_1)$ avoiding 0 in $F$	$1 \times 1$
${}^x\Psi^r(s) = \Psi^r(s) - {}_0\widehat{f}_{22}^r(0, x, s)\Psi^r(s)e^{H^r(s)x}$	from $(0, S_2)$ to $(0, S_1)$ avoiding $x$ in $F^r$	$n \times 1$
${}_0\widehat{h}_{12}(b, b, s) = (sI - Q_{11})^{-1}Q_{12} \times [I - {}^b\Psi^r(s/2)(sI - Q_{11})^{-1}Q_{12}]^{-1}$	from $(b, S_1)$ to $(b, S_1)$ avoiding 0 in ${}^bF$	$1 \times n$
${}^x\widehat{f}_{22}(x, 0, s) = {}_0\widehat{f}_{22}^r(0, x, s)$	from $(x, S_2)$ to $(0, S_2)$ avoiding $x$ in $F$	$n \times n$
${}^x\Psi(s) = \Psi(s) - {}_0\widehat{f}_{11}(0, x, s)\Psi(s)e^{H(s)x}$	from $(0, S_1)$ to $(0, S_2)$ avoiding $x$ in $F$	$1 \times n$
${}_0\widehat{f}_{11}(u, b, s) = [I - {}^{b-u}\Psi(s)u\Psi^r(s)]^{-1} \times {}_0\widehat{f}_{11}(0, b - u, s)$	from $(u, S_1)$ to $(b, S_1)$ avoiding 0 in $F$	$1 \times 1$
${}^b\widehat{f}_{22}(u, 0, s) = [I - {}^u\Psi^r(s)b^{-u}\Psi(s)]^{-1} \times {}^u\widehat{f}_{22}(u, 0, s)$	from $(u, S_2)$ to $(0, S_2)$ avoiding $b$ in $F$	$n \times n$

consider in particular the process

$$X(t) = \sum_{i=1}^{N(t)} V_i - \rho t, \quad t \geq 0$$

with  $N(t)$  and the  $V_i$  as above. In this case, the exponent of  $X(t)$  is

$$\varphi(\alpha) = \alpha\rho + \lambda\pi(\alpha I - G)^{-1} - \lambda.$$

In the next section, we derive the cost functionals for the positive and for the negative inventory process separately. Note that given the phase  $j \in S_2$  at the zero point of a cycle, the two processes are independent in this cycle.

### 4. THE DISCOUNTED MODEL

#### 4.1. The Positive Inventory Process

Recall that  $\tau$  denotes the time until the first zero point. Let  $\hat{\tau}(s)$  be the  $1 \times n$  vector whose  $j$ th component is the restricted LST

$$E(e^{-s\tau} \mathbf{1}_{\{\text{level } 0 \text{ is downcrossed at time } \tau \text{ with phase } j\}}).$$

Ahn and Ramaswami ([4], Theorem 7) have derived an expression for  $\hat{\tau}(s)$  by applying the fluid model and the appropriate changes in the clock time, which we state as Lemma 1.

LEMMA 1:

$$\hat{\tau}(s) = {}^b\Psi(s/2) + e^{\frac{-sb}{2\rho_1}} {}_0\hat{f}_{11}(0, b, s/2) {}_0\hat{h}_{12}(b, b, s) e^{\frac{sb}{2\rho_1}} {}^b\hat{f}_{22}(b, 0, s/2).$$

Now consider a Lévy process  $X_1(t)$  with exponent  $\varphi_1(\alpha) = \alpha\rho_1 + \lambda\pi(\alpha I - G)^{-1} - \lambda$  and  $X_1(0) = 0$ . Let  $L(t) = -\inf_{0 \leq s \leq t} \min[0, b + X_1(s)]$ . The process  $L(t)$  is known as the local time; for  $t < \tau$ , it represents the lost production due to the capacity restriction until time  $t$ . Let  $Z(t) = X_1(t) + L(t)$ . It is not difficult to see that this latter process up to time  $\tau$ , i.e.,  $(Z(t))_{0 \leq t \leq \tau}$ , has the same distribution as  $(-\check{I}(t))_{0 \leq t \leq \tau}$ . Note that  $L(t)$  increases at  $t$  if and only if  $Z(t) = -b$ .

Finally, define  $Y(t) = L(t) + (\beta/\alpha)t$  (for an arbitrary  $\beta \geq 0$ ) and  $W(t) = X_1(t) + Y(t) = Z(t) + (\beta/\alpha)t$ . Since  $Y(t)$  is adapted and has paths of finite expected variation, the process

$$\begin{aligned} M(t, \alpha) &= \varphi_1(\alpha) \int_0^t e^{-\alpha W(s)} ds + e^{-\alpha Y(0)} - e^{-\alpha W(t)} - \alpha \int_0^t e^{-\alpha W(s)} dY(s) \\ &= (\varphi_1(\alpha) - \beta) \int_0^t e^{-\alpha Z(s) - \beta s} ds + 1 - e^{-\alpha Z(t) - \beta t} - \alpha \int_0^t e^{-\alpha Z(s) - \beta s} dL(s) \end{aligned} \tag{17}$$

is a martingale of zero mean (as  $Z(0) = 0$ ). The optional sampling theorem yields  $0 = EM(0, \alpha) = EM(\tau, \alpha)$ , i.e.,

$$E \int_0^\tau e^{-\alpha Z(s) - \beta s} ds = \frac{E(e^{-\alpha Z(\tau) - \beta \tau}) - 1 + \alpha e^{\alpha b} E \int_0^\tau e^{-\beta s} dL(s)}{\varphi_1(\alpha) - \beta}. \tag{18}$$

To apply this equation to our problem we have to find  $E(e^{-\alpha Z(\tau) - \beta \tau})$  and  $E \int_0^\tau e^{-\beta s} dL(s)$ .

LEMMA 2: *We have*

$$E(e^{-\alpha Z(\tau)-\beta\tau}) = \hat{\tau}(\beta)(\alpha I - G)^{-1}G^o,$$

where the vector  $\hat{\tau}(\beta)$  is given by Lemma 1.

PROOF: Given that at the zero point the process hits level 0 at the demand phase  $j \in S_2$ ,  $Z(\tau)$  and  $\tau$  are independent random variables and  $Z(\tau)$  has the phase-type distribution with presentation  $(\epsilon, e_j, G)$ , where  $e_j$  is a unit vector of order  $n$  with 1 in the  $j$ th place and 0 elsewhere. Applying the fluid method and Lemma 1 we obtain

$$\begin{aligned} E(e^{-\alpha Z(\tau)-\beta\tau}) &= \sum_{j=1}^n E(e^{-\alpha Z(\tau)-\beta\tau} \mathbf{1}_{\{\mathcal{J}(\tau)=j\}}) \\ &= \sum_{j=1}^n E(e^{-\beta\tau} \mathbf{1}_{\{\mathcal{J}(\tau)=j\}}) E(e^{-\alpha\epsilon_j}) = \hat{\tau}(\beta)(\alpha I - G)^{-1}G^o. \end{aligned}$$

The lemma is proved. ■

Regarding the term  $E \int_0^\tau e^{-\beta s} dL(s)$ , we note that

$$E \int_0^\tau e^{-\beta s} dL(s) = \rho_1 E \int_0^\tau e^{-\beta s} \mathbf{1}_{\{Z(s)=-b\}} ds = \rho_1 \hat{k}^b(\beta),$$

where  $\hat{k}^b(\beta)$  has been defined as the expected discounted idle time due to the storage maximum  $b$ . Let  $\hat{k}^b(\beta)$  be the  $n$ -column vector whose  $j$ th component is the expected discounted idle time until the zero point, given that  $\check{I}(0) = b, \mathcal{J}(0) = j$ , i.e., given that at time 0 the process exits level  $b$  due to a demand at phase  $j \in S_2$ .

LEMMA 3: *The vector  $\hat{k}^b(\beta)$  is given by*

$$\hat{k}^b(\beta) = \left( I - \frac{\lambda}{\beta + \lambda} {}^b\Psi_1^r \left( \frac{\beta}{2} \right) \pi \right)^{-1} {}^b\Psi_1^r \left( \frac{\beta}{2} \right) \frac{1}{\beta + \lambda}. \tag{19}$$

PROOF: Once the storage level hits  $b$ , the production is stopped for a random time  $\zeta$  until the next demand arrival, where  $\zeta$  is exponentially distributed with rate  $\lambda$ . We can write  $\hat{k}^b(\beta)$  as follows:

$$\hat{k}^b(\beta) = {}^b\Psi_1^r \left( \frac{\beta}{2} \right) E \left( \int_0^\zeta e^{-\beta t} dt \right) + {}^b\Psi_1^r \left( \frac{\beta}{2} \right) E(e^{-\beta\zeta}) \pi \hat{k}^b(\beta). \tag{20}$$

The  $j$ th component of the  $n \times 1$  vector  ${}^b\Psi_1^r(\frac{\beta}{2})$  is the expected discounted time to reach level  $b$  without touching level 0, given that at time 0 the process exits from  $b$  due to a demand at phase  $j$ . The expected discounted duration of the idle period, given that the process reaches level  $b$ , is  $E \left( \int_0^\zeta e^{-\beta t} dt \right)$ . The second term of (20) is the expected discounted time until the second time that the process hits level  $b$ . Solving (20) for  $\hat{k}^b(\beta)$  we obtain (19). ■

LEMMA 4:

$$\hat{k}^b(\beta) = e^{-\beta/2\rho_1} {}_0\hat{f}_{11} \left( 0, b, \frac{\beta}{2} \right) \frac{1}{\beta + \lambda} (1 + \lambda \pi \hat{k}^b(\beta)). \tag{21}$$

PROOF: We can write  $\widehat{k}(\beta)$  as follows:

$$\widehat{k}(\beta) = e^{-\beta/2\rho_1} {}_0\widehat{f}_{11}\left(0, b, \frac{\beta}{2}\right) \left[ E\left(\int_0^\zeta e^{-\beta t} dt\right) + E(e^{-\beta\zeta})\pi\widehat{k}^b(\beta) \right]. \tag{22}$$

According to Table 3 (with the appropriate changes in the clock time),  $e^{-\beta b/2\rho_1} {}_0\widehat{f}_{11}(0, b, \frac{\beta}{2})$  is the expected discounted time to hit level  $b$  first given that  $\check{I}(0) = 0$ . From that point the production is stopped for a time  $\zeta$ . By the proof of Lemma 3, the second term of (22) is the expected discounted idle time after the first exit from level  $b$ . ■

To determine  $-E \int_0^\tau e^{-\beta s} Z(s) ds = E \int_0^\tau e^{-\beta s} \check{I}(s) ds$ , take the derivative on both sides of (18) with respect to  $\alpha$  and let  $\alpha$  tend to 0 in the resulting equation. This yields

$$\widehat{h}(\beta) = E \int_0^\tau e^{-\beta s} \check{I}(s) ds = \frac{-\beta(\widehat{\tau}(\beta)G^{-1}e + \rho_1\widehat{k}(\beta)) - (\widehat{\tau}(\beta)e - 1)(\rho_1 - \lambda E(V))}{\beta^2}. \tag{23}$$

For  $b \rightarrow \infty$  we have  $\widehat{\tau}(\beta) \rightarrow \Psi_1(\frac{\beta}{2})$ ,  $\widehat{k}(\beta) \rightarrow 0$ . Thus, we get for the limiting case of infinite capacity

$$E \int_0^\tau e^{-\beta s} \check{I}(s) ds = \frac{-\beta\left(\Psi_1(\frac{\beta}{2})G^{-1}e\right) - \left(\Psi_1(\frac{\beta}{2})e - 1\right)(\rho_1 - \lambda E(V))}{\beta^2}.$$

### 4.2. The Negative Inventory Process

The negative inventory process stays at level 0 up to time  $\tau$  and then moves in  $[-a, 0)$  until time  $T$ . We shift the time origin to  $\tau$ . Recall that back-ordering up to level  $-a$  is allowed and that any backlog that exceeds this level is lost so that the inventory level thereafter restarts from level  $-a$ .

Let  $X_2(t)$  be a Lévy process starting at  $X_2(0) = 0$  and having exponent

$$\varphi_2(\alpha) = \alpha\rho_2 + \lambda\pi(\alpha I - G)^{-1} - \lambda.$$

Assume that at the zero point the demand phase is  $i \in S_2$ . Let  $\varepsilon_i \sim PH(\epsilon, e_i, G)$ . Let  $Y_i(t) = \varepsilon_i + (\beta/\alpha)t$  (for an arbitrary  $\beta \geq 0$ ).  $Y_i(t)$  is an adapted process and has paths of finite expected variation. Let  $Z_i(t) = X_2(t) + \varepsilon_i$  and  $W_i(t) = X_2(t) + Y_i(t) = Z_i(t) + (\beta/\alpha)t$ .

Now define the stopping times

$$L_i = \inf\{t \geq 0 : Z_i(t) = 0 \text{ or } Z_i(t) > a\}, \tag{24}$$

$$\widetilde{T}_i = \inf\{t \geq 0 : Z_i(t) = 0\}. \tag{25}$$

The distribution of the process  $(Z_i(t))_{0 \leq t < L_i}$  is the same as the conditional distribution of  $(-\check{I}(\tau + t))_{0 \leq t < L_i}$ , given that at the zero point the demand phase is  $i \in S_2$ .

Similarly as above, we introduce  $Y^a(t) = a + (\beta/\alpha)t$ ,  $Z^a(t) = X_2(t) + a$  and  $W^a(t) = X_2(t) + Y^a(t) = Z^a(t) + (\beta/\alpha)t$ . Moreover, we define the stopping times  $L^a, \widetilde{T}^a$  associated to the process  $Z^a(t)$ . Clearly  $Z^a(t)$  describes the reflected negative process for the case of initial inventory level  $a$  due to lost sales.

Let

$$R_i(\alpha) = E_i \int_0^{\tilde{T}_i} e^{-\alpha Z_i(s) - \beta s} ds, \tag{26}$$

$$R^a(\alpha) = E_a \int_0^{\tilde{T}^a} e^{-\alpha Z^a(s) - \beta s} ds, \tag{27}$$

where  $E_i$  ( $E_a$ ) denotes conditional expectation given that the initial phase is  $i \in S_2$  (given that  $Z^a(0) = a$ ). For convenience, we will drop the index  $i$  and write  $L, \tilde{T}, Z(t)$  instead of  $L_i, \tilde{T}_i, Z_i(t)$ . Next, we compute the vector

$$\mathbf{R}(\alpha) = (R_1(\alpha), \dots, R_n(\alpha))^T.$$

PROPOSITION 1:

$$R_i(\alpha) = E_i \int_0^L e^{-\alpha Z(s) - \beta s} ds + E_i [e^{-\beta L} \mathbf{1}_{\{L < \tilde{T}\}}] R^a(\alpha), \tag{28}$$

$$R^a(\alpha) = E \int_0^{L^a} e^{-\alpha Z^a(s) - \beta s} ds + E [e^{-\beta L^a} \mathbf{1}_{\{L^a < \tilde{T}^a\}}] R^a(\alpha). \tag{29}$$

PROOF:

$$\begin{aligned} R_i(\alpha) &= E_i \int_0^{\tilde{T}} e^{-\alpha Z(s) - \beta s} ds = E_i \int_0^L e^{-\alpha Z(s) - \beta s} ds + E_i \int_L^{\tilde{T}} e^{-\alpha Z(s) - \beta s} ds \\ &= E_i \int_0^L e^{-\alpha Z(s) - \beta s} ds + E_i \int_L^{\tilde{T}} e^{-\alpha Z(s) - \beta s} \mathbf{1}_{\{L < \tilde{T}\}} ds + E_i \int_L^{\tilde{T}} e^{-\alpha Z(s) - \beta s} \mathbf{1}_{\{L = \tilde{T}\}} ds \\ &= E_i \int_0^L e^{-\alpha Z(s) - \beta s} ds + E_i \int_0^{\tilde{T}-L} e^{-\alpha Z(u+L) - \beta(u+L)} \mathbf{1}_{\{L < \tilde{T}\}} du. \end{aligned}$$

On the event  $\{L < \tilde{T}\}$  the process  $Z(s)$  exceeds level  $a$  before reaching level 0 and then restarts from  $a$ . In this case,  $Z(L + u)$  is distributed as  $Z^a(u)$ ,  $\tilde{T} - L$  is distributed as  $\tilde{T}^a$ , and we obtain

$$\begin{aligned} R_i(\alpha) &= E_i \int_0^L e^{-\alpha Z(s) - \beta s} ds + E_i (e^{-\beta L} \mathbf{1}_{\{L < \tilde{T}\}}) E \int_0^{\tilde{T}^a} e^{-\alpha Z^a(s) - \beta s} ds \\ &= E_i \int_0^L e^{-\alpha Z(s) - \beta s} ds + E_i [e^{-\beta L} \mathbf{1}_{\{L < \tilde{T}\}}] R^a(\alpha). \end{aligned} \tag{30}$$

The proof of (29) is similar. ■

Rearrange terms in (29) to obtain

$$R^a(\alpha) = \frac{E \int_0^{L^a} e^{-\alpha Z^a(s) - \beta s} ds}{1 - E [e^{-\beta L^a} \mathbf{1}_{\{L^a < \tilde{T}^a\}}]}. \tag{31}$$

Regarding the denominator in (31), note that

$$E [e^{-\beta L^a} \mathbf{1}_{\{L^a < \tilde{T}^a\}}] = {}^a\Psi_2\left(\frac{\beta}{2}\right)\mathbf{1}.$$

Next, we have to determine the numerator in (31), i.e.,  $E \int_0^{L^a} e^{-\alpha Z^a(s) - \beta s} ds$ . We use the Kella–Whitt martingale

$$M^a(\alpha, t) = \varphi_2(\alpha) \int_0^t e^{-\alpha W^a(s)} ds + e^{-\alpha Y^a(0)} - e^{-\alpha W^a(t)} - \alpha \int_0^t e^{-\alpha W^a(s)} dY^a(s) \tag{32}$$

$$= (\varphi_2(\alpha) - \beta) \int_0^t e^{-\alpha Z^a(s) - \beta s} ds + e^{-\alpha a} - e^{-\alpha Z^a(t) - \beta t}. \tag{33}$$

Its mean is zero, as  $Z^a(0) = a$ . Applying the optional sampling theorem yields  $0 = EM^a(\alpha, 0) = EM^a(\alpha, L^a)$ , which is tantamount to

$$(\varphi_2(\alpha) - \beta) E \int_0^{L^a} e^{-\alpha Z^a(s) - \beta s} ds = E(e^{-\alpha Z^a(L^a) - \beta L^a}) - e^{-\alpha a} \tag{34}$$

$$= E(e^{-\alpha Z^a(L^a) - \beta L^a} \mathbf{1}_{\{L^a = \tilde{T}^a\}}) + E(e^{-\alpha Z^a(L^a) - \beta L^a} \mathbf{1}_{\{L^a < \tilde{T}^a\}}) - e^{-\alpha a}. \tag{35}$$

If  $L^a = \tilde{T}^a$  then  $Z^a(L^a) = 0$  while if  $L^a < \tilde{T}^a$ , conditional on the process hitting level  $a$  at phase  $j$ , we have  $Z^a(L^a) = a + \varepsilon_j$ , and  $Z^a(L^a)$  is independent of the  $L^a$  and the past before  $L^a$ . Now the fluid flow approach yields

$$E(e^{-\alpha Z^a(L^a) - \beta L^a} \mathbf{1}_{\{L^a = \tilde{T}^a\}}) = e^{-\beta/2\rho_2} \widehat{f}_{11} \left( 0, a, \frac{\beta}{2} \right), \tag{36}$$

$$E(e^{-\alpha Z^a(L^a) - \beta L^a} \mathbf{1}_{\{L^a < \tilde{T}^a\}}) = e^{-\alpha a} {}^a\Psi_2 \left( \frac{\beta}{2} \right) (\alpha I - G)^{-1} G^o. \tag{37}$$

Substituting (36) and (37) into (34) we find that

$$E \int_0^{L^a} e^{-\alpha Z^a(s) - \beta s} ds = \frac{-e^{-\alpha a} + e^{-\beta/2\rho_2} \widehat{f}_{11} \left( 0, a, \frac{\beta}{2} \right) + e^{-\alpha a} {}^a\Psi_2 \left( \frac{\beta}{2} \right) (\alpha I - G)^{-1} G^o}{\varphi_2(\alpha) - \beta}. \tag{38}$$

Substituting (38) into (31) leads to a formula for  $R^a(\alpha)$ .

Similarly, we can derive the terms  $E_i \int_0^L e^{-\alpha Z(s) - \beta s} ds$  and  $E_i e^{-\beta L} \mathbf{1}_{\{L < \tilde{T}\}}$ , relying on the martingale

$$M_i(\alpha, t) = (\varphi_2(\alpha) - \beta) \int_0^t e^{-\alpha Z_i(s) - \beta s} ds + e^{-\alpha Y_i(0)} - e^{-\alpha Z_i(t) - \beta t}. \tag{39}$$

The process  $Z_i(t)$  starts from phase  $i \in S_2$ . We first apply the fluid description of the process and obtain the vector representations

$$\mathbf{E}(e^{-\alpha Z(L) - \beta L} \mathbf{1}_{\{L = \tilde{T}\}}) = {}^a\Psi_2^r \left( \frac{\beta}{2} \right), \tag{40}$$

$$\mathbf{E}(e^{-\alpha Z(L) - \beta L} \mathbf{1}_{\{L < \tilde{T}\}}) = e^{-\alpha a} e^{\frac{\beta a}{2\rho_2}} \widehat{f}_{22} \left( a, 0, \frac{\beta}{2} \right) (\alpha I - G)^{-1} G^o. \tag{41}$$

Second, we use the optional sampling theorem for (39), and inserting (40) yields the  $(n \times 1)$  expectation vector

$$\mathbf{E} \int_0^L e^{-\alpha Z(s) - \beta s} ds = \frac{-\mathbf{E}e^{-\alpha \varepsilon} + {}^a\Psi_2^r \left( \frac{\beta}{2} \right) + e^{-\alpha a} e^{\frac{\beta a}{2\rho_2}} \widehat{f}_{22} \left( a, 0, \frac{\beta}{2} \right) (\alpha I - G)^{-1} G^o}{\varphi_2(\alpha) - \beta}. \tag{42}$$

Here,  $\mathbf{E}e^{-\alpha \varepsilon}$  is an  $(n \times 1)$  vector. It is equal to  $(\alpha I - G)^{-1} G^o$ .

From (42) and (31) we obtain a formula for the vector  $\mathbf{R}(\alpha)$ .

Finally, to determine  $\widehat{s}_i(\beta) = -E_i \int_0^{\widetilde{T}} e^{-\beta s} \check{I}(s + \tau) ds$ , we take the derivative of (26) with respect to  $\alpha$  and let  $\alpha$  tend to 0.

REMARK: For the limiting case  $a = \infty$  (no backlog restriction), we get

$$R_i(\alpha) = E_i \int_0^{\widetilde{T}} e^{-\alpha Z(s) - \beta s} ds = \frac{E_i[e^{-\alpha Z(\widetilde{T}) - \beta \widetilde{T}}] - E_i(e^{-\alpha \varepsilon_i})}{\varphi_2(\alpha) - \beta}. \tag{43}$$

Note that  $Z(\widetilde{T}) = 0$ ,  $E_i(e^{-\beta \widetilde{T}}) = e_i \Psi_2^r(\frac{\beta}{2})$ ,  $E_i(e^{-\alpha \varepsilon_i}) = e_i(\alpha I - G)^{-1} G^o$ . Substituting in (43), taking the derivative with respect to  $\alpha$  and then letting  $\alpha \rightarrow 0$ , we obtain the Laplace transform vector

$$\widehat{\mathbf{s}}(\beta) = -\beta^{-2} \left[ \beta G^{-1} e - \left( \Psi_2^r \left( \frac{\beta}{2} \right) - e \right) (\rho_2 - \lambda E(V)) \right]. \tag{44}$$

Next, we turn to the cost of unsatisfied demand. Recall that we denote the occurrence time of the  $k$ th unsatisfied demand by  $\xi_k$  and the lost amount by  $Y_k$ .

Let  $\widehat{u}_a(\beta)$  be the  $(n \times 1)$  vector whose  $i$ th component is the conditional expected discounted total value of the unsatisfied demands, given that the negative process starts at phase  $i \in S_2$  and at time 0. Denote by  $u_0(\beta)$ , the conditional expected discounted total value of unsatisfied demands, given the process starts at level  $-a$  (a scalar).

PROPOSITION 2: *The vector  $\widehat{u}_a(\beta)$  and the scalar  $\widehat{u}_0(\beta)$  are given by*

$$\widehat{u}_a(\beta) = e^{\beta a / 2 \rho_2} {}^a \widehat{f}_{22} \left( a, 0, \frac{\beta}{2} \right) [(-G)^{-1} e + e \widehat{u}_0(\beta)], \tag{45}$$

$$\widehat{u}_0(\beta) = {}^a \Psi_2 \left( \frac{\beta}{2} \right) [(-G)^{-1} e + e \widehat{u}_0(\beta)]. \tag{46}$$

PROOF: Let us change the  $Y$ -coordinate such that level 0 becomes level  $a$ , and level  $-a$  becomes 0. Thus, the negative process starts at level  $a$ . If the process hits level 0 before level  $a$ , say at phase  $j$ , an unsatisfied demand occurs. By applying the fluid model and the appropriate changes in the clock time, the LST of that time is  $e^{\beta a / 2 \rho_2} e_j {}^a \widehat{f}_{22}(a, 0, \frac{\beta}{2})$ . The size of the associated unsatisfied demand,  $\varepsilon_j$ , has a  $PH(\epsilon, e_j, G)$  distribution. Then the process restarts from level 0 and its expected discounted demand loss from then on is  $\widehat{u}_0(\beta)$ . This yields (45) (in a vector form). From (45) we get  $\widehat{u}_0(\beta)$  by setting  $a = 0$ . ■

To derive  $\widehat{u}(\beta)$  we use Proposition 2 and the LST of  $\tau$ . We obtain

$$\widehat{u}(\beta) = \mathbf{E}(e^{-\beta \tau}) \widehat{u}_a(\beta) = \widehat{\tau}(\beta) \widehat{u}_a(\beta). \tag{47}$$

### 4.3. The LST of a Cycle Length

To complete the treatment of the discounted model, we finally need to determine the LST of a cycle length. Clearly,

$$E(e^{-\beta T}) = E[e^{-\beta(\tau + T - \tau)}] = E[e^{-\beta \tau} e^{-\beta \widetilde{T}}].$$

Given that the phase at the zero point is  $j \in S_2$ , we have that  $e^{-\beta \tau}$  and  $e^{-\beta \widetilde{T}}$  are independent. Thus,

$$E(e^{-\beta T}) = \mathbf{E}[e^{-\beta \tau}] \mathbf{E}[e^{-\beta \widetilde{T}}] = \widehat{\tau}(\beta) \mathbf{E}[e^{-\beta \widetilde{T}}].$$

To derive the LST of the negative process cycle, let the scalar  $E_0(\beta)$  be the LST of the time until the process exceeds level  $(-a)$  before level 0, given that the process starts at level  $-a$ .

PROPOSITION 3:  $\mathbf{E}e^{-\beta\tilde{T}}$  and  $E_0(\beta)$  are given by

$$\mathbf{E}[e^{-\beta\tilde{T}}] = e^{\frac{\beta a}{2\rho_2}} {}^a\hat{f}_{22}\left(a, 0, \frac{\beta}{2}\right) eE_0(\beta) + {}^a\Psi_2^r\left(\frac{\beta}{2}\right), \tag{48}$$

$$E_0(\beta) = {}^a\Psi_2\left(\frac{\beta}{2}\right) eE_0(\beta) + e^{\frac{-\beta a}{2\rho_2}} {}_0\hat{f}_{11}\left(0, a, \frac{\beta}{2}\right). \tag{49}$$

PROOF: Assume that the negative process starts with phase  $i \in S_2$ . We apply the fluid model and the appropriate changes in the clock time. There are two possibilities: either the process hits level  $(-a)$  before level 0; or the LST of this time is  $e^{\frac{\beta a}{2\rho_2}} e_i {}^a\hat{f}_{22}(a, 0, \frac{\beta}{2})e$ . The second possibility is to reach the recovery point without exceed level  $(-a)$ . The LST of this time is  $e_i {}^a\Psi_2^r(\frac{\beta}{2})$ . This yields the vector representation (48). The scalar  $E_0(\beta)$  is obtained similarly. ■

REMARK: For  $a \rightarrow \infty$  we get  $\mathbf{E}e^{-\beta\tilde{T}} \rightarrow \Psi_2^r(\frac{\beta}{2})$  and  $\mathbf{E}e^{-\beta\tau} \rightarrow \Psi_1(\frac{\beta}{2})$ .

### 5. NUMERICAL EXAMPLE

Example 1: We consider an underlying Poisson process with intensity  $\lambda = 2$ . The demand size has a phase-type distribution with initial probability vector  $\pi = (0.5614, 0.4386)$  and a transition rate matrix

$$G = \begin{pmatrix} -8.64 & 1.997 \\ 0.101 & -1.095 \end{pmatrix}.$$

We fix the discount factor, the storage capacity and the threshold level as follows:

$$\beta = 0.01, \quad b = 5, \quad a = 7.$$

Let  $\rho_1$  vary in  $\{0.2, 0.3, 0.4, 0.5, 0.6\}$  and  $\rho_2$  vary in  $\{1, 1.25, 1.5, 1.75, 2\}$ . We determine the expected discounted holding cost, shortage cost and unsatisfied cost components. We do not reproduce lengthy tables of calculated values (which are not very illuminating) but instead present our results in terms of graphical displays. Figures 2, 3 and 4 show  $HC(\beta)/h, SC(\beta)/w$  and  $UC(\beta)/\varphi$  as functions of  $\rho_1$  and  $\rho_2$ , respectively. (We could also have simply set  $h, w$  and  $\varphi$  equal to 1.)

In Figure 2, we see that the holding cost increases as a function of  $\rho_2$  and appears to be a concave function of  $\rho_1$ . Moreover, we can conclude that for each value of  $\rho_2$  there is a maximum holding cost, which increases as  $\rho_2$  increases.

Figures 3 and 4 look similar, showing  $SC(\beta)/w$  and  $UC(\beta)/\varphi$  as decreasing functions of  $\rho_1$  and  $\rho_2$ . They are convex in  $\rho_2$  for fixed  $\rho_1$ .

### 6. THE LONG-RUN AVERAGE COST

Let us finally deal with the various long-run average cost criterion. For the positive inventory process, we derive the holding cost  $\hat{h}(0)$  and the idle cost due to the finite storage capacity  $b$ .



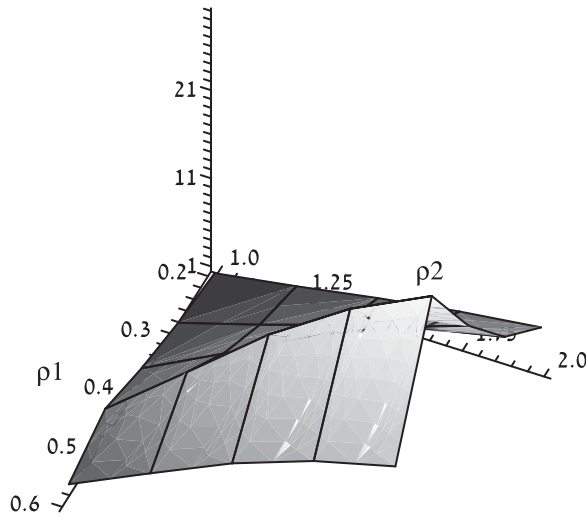


FIGURE 2. The holding cost component,  $HC(\beta)/h$ .

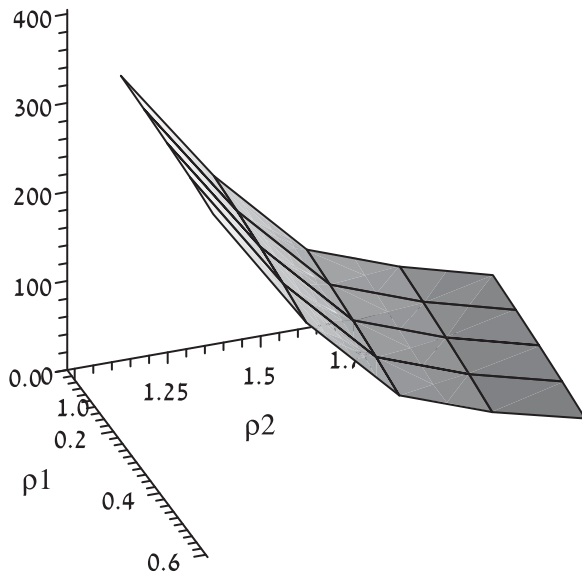


FIGURE 3. The shortage cost component,  $SC(\beta)/w$ .

Setting  $\beta = 0$  in (18) yields

$$E \int_0^\tau e^{-\alpha Z(s)} ds = \frac{\hat{\tau}(0)(\alpha I - G)^{-1}G^o - 1 + \alpha e^{\alpha b}EL(\tau)}{\varphi_1(\alpha)}. \tag{50}$$

$\hat{\tau}(0)$  is an  $1 \times n$  probability row vector, where  $\hat{\tau}(0)_j$  is the probability that the phase at the zero point is  $j \in S_2$ . Take the derivative with respect to  $\alpha$ , let  $\alpha$  tend to 0 and use

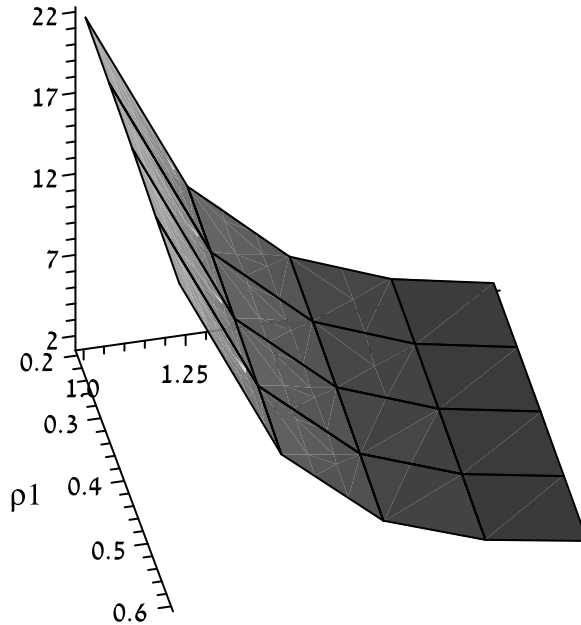


FIGURE 4. The unsatisfied cost component,  $UC(\beta)/\varphi$ .

l'Hôpital's rule. One finds that

$$\hat{h}(0) = E \int_0^\tau \check{I}(s) ds = \frac{[E(V_\tau^2) + 2bEL(\tau)] (\rho_1 - \lambda E(V)) + \lambda E(V^2) [E(V_\tau) - EL(\tau)]}{2(\rho_1 - \lambda E(V))^2}, \tag{51}$$

where  $V_\tau$  has a phase type-distribution with representation  $PH(\epsilon, \hat{\tau}(0), G)$ .

To complete the derivation of  $\hat{h}(0)$  we have to find  $EL(\tau)$ ; note that  $EL(\tau) = \rho_1 \hat{k}(0)$ .

Dickson and Waters [7] derived the moment-generating function of  $L(\tau)$  as a function of  $q_0$  and  $q_b$ , where  $q_0$  ( $q_b$ ) is the probability of hitting level  $b$  before the zero point, given that the initial inventory is 0 ( $b$ ). They showed that

$$EL(\tau) = \frac{\rho_1 q_u}{\lambda(1 - q_b)}.$$

Thus, the total amount of production loss until the zero point is 0 with probability  $1 - q_0$ , and with probability  $q_0$  it is exponentially distributed with parameter  $(1 - q_b)\lambda/\rho_1$ . In our case,  $q_0 = 0 \hat{f}_{11}(0, b, 0)$  and  $q_b = \pi^b \Psi_1^r(0)$ . Hence, we obtain

$$\hat{k}(0) = \frac{0 \hat{f}_{11}(0, b, 0)}{\lambda(1 - \pi^b \Psi_1^r(0))}. \tag{52}$$

REMARK: For  $b \rightarrow \infty$  we have  $\hat{\tau}(0) \rightarrow \Psi_1(0)$ . From (51) and  $EL(\tau) = 0$ , it follows that in the case of infinite storage capacity

$$\hat{h}(0) = \frac{E(V_\Psi^2) (\rho_1 - \lambda E(V)) + E(V_\Psi) \lambda E(V^2)}{2 (\rho_1 - \lambda E(V))^2},$$

where  $V_\Psi$  has a phase-type distribution with representation  $PH(\epsilon, \Psi_1(0), G)$ .

For the *negative inventory process*, we derive two costs:  $\hat{s}(0)$ , the shortage cost for backlogged orders, and  $\hat{u}(0)$ , the lost sales cost due to the minimum level  $-a$ .

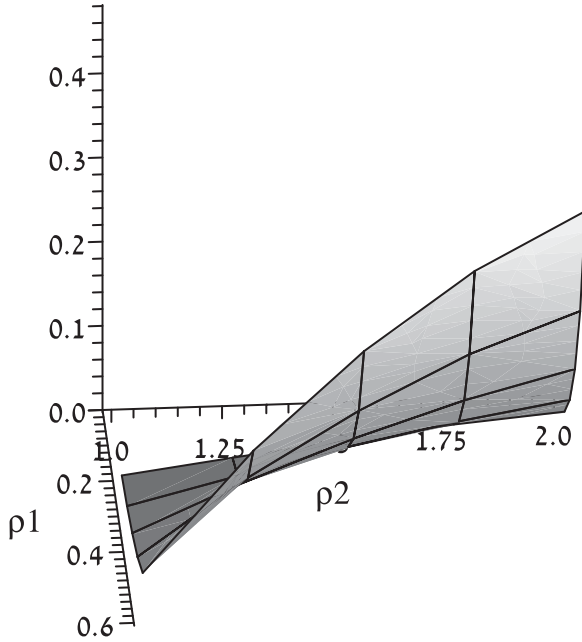


FIGURE 5. The expected value of the inventory level.

Assume that the process starts at phase  $i \in S_2$ . Inserting  $\beta = 0$  in (26) and (27) yields

$$\begin{aligned}
 R_i(\alpha) &= E_i \int_0^L e^{-\alpha Z(s)} ds + P_i(L < \tilde{T}) R^a(\alpha), \\
 R^a(\alpha) &= \frac{E_a \int_0^{L^a} e^{-\alpha Z^a(s)} ds}{1 - P_a(L^a < \tilde{T}^a)} = \frac{-e^{-\alpha a} + (1 - {}^a\Psi_2(0)e) + (e^{-\alpha a}) {}^a\Psi_2(0)(\alpha I - G)^{-1} G^o}{\varphi_2(\alpha)[1 - {}^a\Psi_2(0)e]}.
 \end{aligned}
 \tag{53}$$

Using the fluid method, we obtain the  $(n \times 1)$  probability vector  $P_i(L < \tilde{T}) = {}^a\hat{f}_{22}(a, 0, 0)e$ . Inserting  $\beta = 0$  in (42) leads to the  $(n \times 1)$  vector

$$\mathbf{R}(\alpha) = \frac{-(\alpha I - G)^{-1} G^o + {}^a\Psi_2^r(0) + e^{-\alpha a} {}^o\hat{f}_{22}(a, 0, 0)(\alpha I - G)^{-1} G^o}{\varphi_2(\alpha)} + {}^a\hat{f}_{22}(a, 0, 0) R^a(\alpha).
 \tag{54}$$

Now take the derivative with respect to  $\alpha$ , let  $\alpha$  tend to 0 and use l'Hôpital's rule twice to get  $\hat{s}(0)$ .

REMARK: Consider the case  $a = \infty$ . Note that  $\hat{\tau}(0)$  is the row vector of the probabilities of the phases at time  $\tau$ . Now inserting  $\beta = 0$  in (43), taking the derivative with respect to  $\alpha$ , letting  $\alpha$  tend to 0 and using l'Hôpital's rule, we arrive at

$$\hat{s}(0) = \frac{E(V_\tau^2) (\rho_2 - \lambda E(V)) + E(V_\tau) \lambda E(V^2)}{2 (\rho_2 - \lambda E(V))^2}.$$

Of course, as  $a = \infty$  is the case of arbitrary backlog, we have  $\hat{u}(0) = 0$ .

For the long-average cost analysis, we also need to know the expected cycle time  $E(T)$ . Recall that  $E(T) = E(\tau) + E(\tilde{T})$ .

Take (50), set  $\alpha = 0$ , use l'Hôpital's rule and obtain

$$E(\tau) = \frac{EL(\tau) - E(V_\tau)}{\rho_1 - \lambda E(V)}. \quad (55)$$

$EL(\tau)$  can be found in (52). For  $b \rightarrow \infty$  we obtain  $E(\tau) = \Psi_1(0)G^{-1}e / (\rho_1 - \lambda E(V))$ .

Regarding  $E(\tilde{T})$ , let  $\alpha$  tend to 0 in (53), use l'Hôpital's rule to get a formula for the vector  $\mathbf{R}(0)$ :

$$\mathbf{R}(0) = \frac{-a {}^a \hat{f}_{22}(a, 0, 0)e - {}^a \hat{f}_{22}(a, 0, 0)(-G^{-1})e - G^{-1}e}{\rho_2 - \lambda E(V)}.$$

Thus,

$$E(\tilde{T}) = \hat{\tau}(0)\mathbf{R}(0). \quad (56)$$

In the case  $a = \infty$ , we obtain  $E(\tilde{T}) = -\hat{\tau}(0)G^{-1}e / (\rho_2 - \lambda E(V))$ .

*Example 2:* We consider the same data as in the Example section above. Figure 5 displays the expected values of the inventory level in steady state. As expected,  $E(H_c)$  increases as a function of  $\rho_1$  and  $\rho_2$ .

#### Acknowledgement

This work was supported by the Deutsche Forschungsgemeinschaft (grant no. 306/13-2).

#### References

1. Ahn, S. & Ramaswami, V. (2004). Transient analysis of fluid flow Models via stochastic coupling to a queue. *Stochastic Models* 20: 71–101.
2. Ahn, S. & Ramaswami, V. (2005). Efficient algorithms for transient analysis of stochastic fluid flow models. *Journal of Applied Probability* 42: 531–549.
3. Ahn, S. & Ramaswami, V. (2006). Transient analysis of fluid models via elementary level-crossing arguments. *Stochastic Models* 22: 129–147.
4. Ahn, S., Badescu, A.L. & Ramaswami, V. (2007). Time dependent analysis of finite buffer fluid flows and risk models with a dividend barrier. *Queueing Systems* 55: 207–222.
5. Baek, J.W., Lee, H.W., Lee, S.W. & Ahn, S. (2011). A Markov modulated fluid flow queueing model under D-policy. *Numerical Linear Algebra with Applications* 18: 993–1010.
6. Boxma, O.J., Perry, D. & Stadje, W. (2001). Clearing models for M/G/1 queues. *Queueing Systems* 38: 287–306.
7. Dickson, D.C.M. & Waters, H.R. (2004). Some optimal dividends problems. *ASTIN Bulletin* 34: 49–74.
8. Doshi, B.T., Van Der Duyn Schouten, F.A. & Talman, J.J. (1978). A Production inventory control model with a mixture of back-orders and lost-sales. *Management Science* 24: 1078–1087.
9. Kella, O. & Whitt, W. (1992). Useful martingales for stochastic storage processes with Lévy input. *Journal of Applied Probability* 29: 396–403.
10. Kella, O., Perry, D. & Stadje, W. (2003). A stochastic clearing model with Brownian and a compound Poisson component. *Probability in the Engineering and Informational Sciences* 17: 1–22.
11. Kulkarni, V.G. & Yan, K. (2007). A fluid model with upward jumps at the boundary. *Queueing Systems* 56: 103–117.
12. Kulkarni, V.G. & Yan, K. (2012). Production-inventory systems in stochastic environment and stochastic lead times. *Queueing Systems* 70: 207–231.
13. Perry, D., Berg, M. & Posner, M.J.M. (2001). Stochastic models for broker inventory in dealership markets with a cash management interpretation. *Insurance: Mathematics and Economics* 29: 23–34.

14. Perry, D., Stadjé, W. & Zacks, S. (2005). Sporadic and continuous clearing policies for a production/inventory system under an  $M/G$  demand process. *Mathematics of Operations Research* 30: 354–368.
15. Ramaswami, V. (2006). Passage times in fluid models with application to risk processes. *Methodology and Computing in Applied Probability* 8: 497–515.
16. Yan, K. & Kulkarni, V.G. (2008). Optimal inventory policies under stochastic production and demand rates. *Stochastic Models* 24: 173–190.