A TWO-QUEUE MODEL WITH ALTERNATING LIMITED SERVICE AND STATE-DEPENDENT SETUPS

E.M.M. Winands[1,2]**, I.J.B.F. Adan**[1] **and G.J. van Houtum**[2]
[1]*Department of Mathematics and Computer Science,* [2]*Department of Technology Management*
Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands
$\{$*i.j.b.f.adan,g.j.v.houtum,e.m.m.winands*$\}$*@tue.nl*

Abstract: We consider a two-queue model with state-dependent setups, in which a single server alternately serves the two queues. The high-priority queue is served exhaustively, whereas the low-priority queue is served according to the k-limited strategy. We obtain the transforms of the queue length and sojourn time distributions under the assumption of Poisson arrivals, generally distributed service times and generally distributed setup times. It is shown how the results of this analysis can be applied in the evaluation of a stochastic two-item single-capacity production system.

Keywords: polling model, two queues, k-limited and exhaustive service policy, state-dependent setups

1 Introduction

The present paper considers a two-queue model with state-dependent setups, i.e., a setup is incurred for a queue only when it is non-empty, in which the single server serves the high-priority queue *exhaustively* and the low-priority queue according to the *k-limited* service strategy. Under this k-limited strategy the server continues working at a queue until either a predefined number of k customers is served or until the queue becomes empty, whichever occurs first.

The motivation for the present study is two-fold. The first one is application-oriented. In particular, in many stochastic multi-product single-capacity make-to-stock production systems considerable setup times are incurred, i.e., the so-called *stochastic economic lot scheduling problem* (SELSP). The presence of these setup times in combination with the stochastic environment are the key complicating factors of the SELSP. On the one hand, one seeks for short cycle lengths, and thus frequent production opportunities for the various products, in order to be able to react to the stochasticity in the system. On the other hand, short cycle lengths will increase the setup frequency, which has a negative influence on the amount of capacity available for production. Consequently, this effect will hinder the timely fulfillment of demand. In the context of the SELSP, the exhaustive service discipline has been studied by Federguen and Katalan [5]. A major drawback of this exhaustive policy is that one single product, for which a high demand arrives in a certain period of time, may occupy the machine for quite a while. The impacts of this phenomenon on the other products are stock outs, highly variable cycle lengths and high costs. The k-limited policy circumvents this drawback and offers the possibility to the manager to control both the setup frequencies and the cycle lengths.

A second motivation for the present work is the fact that, so far, hardly any exact results for multi-queue systems with the k-limited service policy have been obtained; even mean performance measures are, in general, not known (see, e.g., Resing [11]). For general k, an exact evaluation is only available for very few special two-queue cases (see, e.g., Lee [8] and Ozawa [9, 10]). In particular, Lee [8] studies a model similar to the one of the present paper under the assumption of *zero setup times*. In many applications, however, the setup times may be substantial and the presence of these setup times may be crucial for the operation of the system.

The contribution of the present paper is two-fold. First, the model in [8] is generalized by including state-dependent setups. In particular, we obtain the transforms of the queue length and sojourn time distributions under the assumption of Poisson arrivals, generally distributed service times and generally distributed setup times. Secondly, we demonstrate how the results of the analysis can be applied in the evaluation of a stochastic two-item single-capacity production system.

The rest of the present paper is organized as follows. In Section 2, we present the model description including the stability conditions and the balance equations. Section 3 derives the *probability generating*

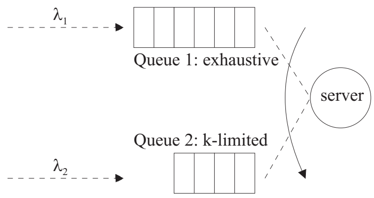

Figure 1: The model.

functions (PGFs) of the joint and marginal queue length distributions both at service completions epochs and at arbitrary instants. Finally, Section 4 is devoted to an application of the analysis.

2 Model description

We start this section with the description of the notation and assumptions used throughout the present paper. Then, Subsections 2.2 and 2.3 present the stability condition and the state description together with the corresponding balance equations, respectively.

2.1 Notation and assumptions

We study a two-queue model, in which a single server alternately serves the two queues. The high-priority queue 1 is served *exhaustively,* whereas the low-priority queue 2 is served according to the *k-limited* strategy (see Figure 1). Customers arrive at queue i according to a Poisson process with rate $\lambda_i > 0$. The service times at queue i are independent, identically distributed random variables with mean $\beta_i > 0$ and *Laplace Stieltjes Transform* (LST) $B_i(\cdot)$. When the server starts service at queue i, a setup time is incurred with mean $\tau_i \geq 0$ and LST $T_i(\cdot)$. These setup times are identically distributed random variables independent of any other event involved. In particular, they are independent of the service times. The occupation rate ρ_i at queue i is defined by $\rho_i = \lambda_i \beta_i$ and the total occupation rate ρ is given by $\rho = \rho_1 + \rho_2$. In the next paragraph, we give a stability condition for the system in terms of the total occupation rate and the service parameter $k \in \{1, 2, \ldots\}$.

The setup times are assumed to be *state-dependent*, i.e., the server incurs a setup for a queue only when it is non-empty. When both queues are empty, the server stops working. He starts again upon arrival of a new customer and, then, he has to setup irrespective of the type of the last customer served before the idle time. Moreover, if the server has served k customers of the low-priority queue and the high-priority queue is empty, the server starts a new sequence up to k customers of this low-priority class after a new setup time.

Finally, we define $S_i(z_1, z_2)$ and $R_i(z_1, z_2)$ as the PGFs of the number of type-1 and type-2 arrivals during a service time and a setup time at queue i, respectively. That is,

$$S_i(z_1, z_2) = B_i(\lambda_1(1 - z_1) + \lambda_2(1 - z_2)), \qquad i = 1, 2, \tag{1}$$
$$R_i(z_1, z_2) = T_i(\lambda_1(1 - z_1) + \lambda_2(1 - z_2)), \qquad i = 1, 2. \tag{2}$$

The quantities $r_1 = \frac{\lambda_1}{\lambda_1 + \lambda_2}$ and $r_2 = \frac{\lambda_2}{\lambda_1 + \lambda_2}$ denote the probabilities that the server switches to queue 1 and 2 after an idle period, respectively.

2.2 Stability conditions

In this subsection, we derive the stability conditions for the two queues by deploying arguments similar to those used by Ibe and Cheng [6]. Since the k-limited service policy bounds the time the server spends at queue 2, queue 2 will not affect the stability of queue 1. Hence, the stability condition of the latter simply reads $\rho_1 < 1$. If this condition is not satisfied, queue 2 is unstable as well: the server will serve queue 1 for an infinite amount of time and, in the meantime, the queue length of queue 2 grows to infinity.

For queue 2, we define a cycle time as the time between two successive setups of the server at this queue. In a long period of time T, the number of customers arriving at queue 2 equals $\lambda_2 T$. Each customer brings, on average, β_1 units of work into the system. We approximately need $\frac{\lambda_2 T}{k}$ cycles to serve these customers. In each cycle, the server always has to setup for queue 2 and has to setup for queue 1 in case queue 1 is non-empty at the instant that the server has served k customers at queue 2. The latter occurs with the probability q that the number of type-1 arrivals during a setup time and k successive service times at queue 2 is not equal to zero, i.e., $q = 1 - R_2(0,1)(S_2(0,1))^k$.

Hence, the mean total setup time in each cycle equals $\tau_2 + q\tau_1$. Moreover, a fraction ρ_1 of time is used to serve customers at queue 1. Since queue 2 is stable if the server manages to serve all customers arrived in T in an amount of time less than T, we have

$$\rho_1 T + \rho_2 T + \frac{\lambda_2 T}{k}(\tau_2 + q\tau_1) < T, \tag{3}$$

which is equivalent to

$$k > \frac{\lambda_2(q\tau_1 + \tau_2)}{1-\rho}. \tag{4}$$

When queue 2 satisfies (4), ρ_1 is also smaller than 1 and, thus, in this case both queues are stable. In the sequel, we assume that (4) is satisfied, as we restrict ourselves to steady-state behavior.

Remark 2.1. *Notice that the form of (4) is identical to the feasibility condition that arises for the production lengths in the deterministic economic lot scheduling problem.* □

2.3 State description and balance equations

We study the system at imbedded epochs of service completions of customers. The state of the system $\mathbf{Q}(n)$ just after the n^{th} departure from the system can be described by the following three variables:

(i) $Q_1(n)$: the number of customers in queue 1;

(ii) $Q_2(n)$: the number of customers in queue 2;

(iii) $C(n)$: equals zero when the n^{th} departure is a type-1 customer, while it equals the number of type-2 departures since the last setup when the n^{th} departure is a type-2 customer.

The associated stochastic process $\mathbf{Q}(n) = \{(Q_1(n), Q_2(n), C(n)), n = 1, 2, \ldots\}$ is an aperiodic and irreducible three-dimensional Markov chain. Let $\pi(q_1, q_2, c) = \lim_{n\to\infty} P[(Q_1(n), Q_2(n), C(n)) = (q_1, q_2, c)]$ be the equilibrium state probability and define the corresponding PGFs for this Markov chain as follows

$$p_1(z_1, z_2) = \sum_{q_1=0}^{\infty}\sum_{q_2=0}^{\infty} \pi(q_1, q_2, 0) z_1^{q_1} z_2^{q_2}, \tag{5}$$

$$p_{2,j}(z_1, z_2) = \sum_{q_1=0}^{\infty}\sum_{q_2=0}^{\infty} \pi(q_1, q_2, j) z_1^{q_1} z_2^{q_2}, \qquad 1 \le j \le k. \tag{6}$$

Now, the following set of $k+1$ balance equations for equally many unknowns $p_1(z_1, z_2)$ and $p_{2,j}(z_1, z_2)$, $j = 1, 2, \ldots, k$, holds

$$\begin{aligned} p_1(z_1, z_2) &= \frac{S_1(z_1, z_2)}{z_1}\Big\{p_1(z_1, z_2) - p_1(0, z_2) + \\ &\quad \Big[c_0 r_1 z_1 + \sum_{j=1}^{k-1}[p_{2,j}(z_1, 0) - p_{2,j}(0, 0)] + p_{2,k}(z_1, z_2) - p_{2,k}(0, z_2)\Big] R_1(z_1, z_2)\Big\}, \end{aligned} \tag{7}$$

$$p_{2,1}(z_1, z_2) = \frac{S_2(z_1, z_2) R_2(z_1, z_2)}{z_2}\Big\{c_0 r_2 z_2 + \alpha(z_2)\Big\}, \tag{8}$$

$$p_{2,j}(z_1, z_2) = \frac{S_2(z_1, z_2)}{z_2}\Big\{p_{2,j-1}(z_1, z_2) - p_{2,j-1}(z_1, 0)\Big\}, \qquad 2 \le j \le k, \tag{9}$$

with $c_0 = p_1(0,0) + \sum_{j=1}^{k} p_{2,j}(0,0)$ the probability that the system is left idle at a departure epoch and

$$\alpha(z_2) = p_1(0,z_2) - p_1(0,0) + p_{2,k}(0,z_2) - p_{2,k}(0,0). \tag{10}$$

These balance equations are formulated by considering all the possible states at the previous departure epoch that can reach the current state. We explain (7), which describes the case that the current departure is a type-1 customer. First of all, the previous departure could be a type-1 departure which did not leave the system idle. This event corresponds to the term $p_1(z_1,z_2) - p_1(0,z_2)$. Secondly, the term $c_0 r_1 z_1$ represents the event that the previous departure left the system idle and that the first new arriving customer is of type 1. Thirdly, the term $\sum_{j=1}^{k-1}[p_{2,j}(z_1,0) - p_{2,j}(0,0)]$ denotes the event that the last departure was a type-2 customer that was not the k^{th} in the sequence and that left queue 2, but not the complete system, idle. Finally, $p_{2,k}(z_1,z_2) - p_{2,k}(0,z_2)$ corresponds with the event that the last departure was a type-2 customer that was the k^{th} in the sequence and that queue 1 was not empty. The explanations of (8) and (9) are similar.

3 Queue length distributions

In Subsection 3.1, we present the derivation of the PGFs of the joint queue length distributions at service completion epochs, which is a generalization of the method used by Lee [8] for the model without setup times. In Subsection 3.2, these results are used to derive expressions for the PGFs of the marginal queue size distributions at arbitrary instants.

3.1 Joint queue lengths at service completion epochs

To derive the PGFs of the joint queue length distributions at service completion epochs, we successively substitute (9) into itself and, then, into (8) which yields, for $j = 1, 2, \ldots, k$,

$$p_{2,j}(z_1,z_2) = \frac{S_2^j(z_1,z_2)R_2(z_1,z_2)\Big\{c_0 r_2 z_2 + \alpha(z_2)\Big\} - \sum_{l=1}^{j-1} z_2^{j-l} S_2^l(z_1,z_2) p_{2,j-l}(z_1,0)}{z_2^j}. \tag{11}$$

Notice that (11) gives an expression of $p_{2,j}(\cdot,\cdot)$, $j = 1, 2, \ldots, k$, as a function of the unknown functions $\alpha(\cdot)$ and $p_{2,l}(\cdot,0)$, $l = 1, 2, \ldots, j-1$.

Now, we turn our attention to $p_1(\cdot,\cdot)$. Substituting (11) for $j = k$ into (7) gives us, after some straightforward manipulations,

$$\begin{aligned}\Big(z_1 - S_1(z_1,z_2)\Big)p_1(z_1,z_2) \quad = \quad & S_1(z_1,z_2)\Big\{c_0 r_1 z_1 R_1(z_1,z_2) + (R_1(z_1,z_2)-1)p_1(0,z_2) + \\ & \Big(\Big(\frac{S_2(z_1,z_2)}{z_2}\Big)^k R_2(z_1,z_2)\Big[c_0 r_2 z_2 + \alpha(z_2)\Big] - \alpha(z_2) + \\ & \sum_{j=1}^{k-1}\Big[1 - \Big(\frac{S_2(z_1,z_2)}{z_2}\Big)^j\Big]p_{2,k-j}(z_1,0) - c_0\Big)R_1(z_1,z_2)\Big\}.\end{aligned} \tag{12}$$

We eliminate $p_1(0,z_2)$ from the above equation by rewriting (10) as follows

$$\begin{aligned} p_1(0,z_2) \quad &= \quad \alpha(z_2) + p_1(0,0) - p_{2,k}(0,z_2) + p_{2,k}(0,0) \qquad\qquad (13)\\ &= \quad \alpha(z_2) + c_0 - \Big(\frac{S_2(0,z_2)}{z_2}\Big)^k R_2(0,z_2)\Big\{c_0 r_2 z_2 + \alpha(z_2)\Big\} - \sum_{j=1}^{k-1}\Big[1 - \Big(\frac{S_2(0,z_2)}{z_2}\Big)^j\Big]p_{2,k-j}(0,0),\end{aligned}$$

which yields

$$\begin{aligned}\Big(z_1 - S_1(z_1,z_2)\Big)p_1(z_1,z_2) \quad = \quad & S_1(z_1,z_2)\Big\{\Big(\frac{\beta(z_1,z_2)}{z_2^k} - 1\Big)\alpha(z_2) + R_1(z_1,z_2)\sum_{j=1}^{k-1}\frac{\gamma_j(z_1,z_2)}{z_2^k}p_{2,k-j}(z_1,0) \\ & + D(z_1,z_2) - (R_1(z_1,z_2)-1)\sum_{j=1}^{k-1}\frac{\gamma_j(0,z_2)}{z_2^k}p_{2,k-j}(0,0)\Big\},\end{aligned} \tag{14}$$

where

$$D(z_1,z_2) = c_0\Big[z_1R_1(z_1,z_2)+r_2z_2\frac{\beta(z_1,z_2)}{z_2^k}-1\Big], \tag{15}$$

$$\beta(z_1,z_2) = S_2^k(z_1,z_2)R_1(z_1,z_2)R_2(z_1,z_2)-S_2^k(0,z_2)(R_1(z_1,z_2)-1)R_2(0,z_2), \tag{16}$$

$$\gamma_j(z_1,z_2) = z_2^k-z_2^{k-j}S_2^j(z_1,z_2). \tag{17}$$

It is again important to notice that via (14), $p_1(\cdot,\cdot)$ is also expressed as a function of the unknown functions $\alpha(\cdot)$ and $p_{2,j}(\cdot,0)$, $j=1,2,\ldots,k-1$.

It is well-known that the term $z_1-S_1(z_1,z_2)$ in (14) has exactly one zero $z_1=\xi(z_2)$ for each $|z_2|\le 1$ if $\rho_1<1$. More specifically,

$$z_1=\xi(z_2)=\gamma_1[\lambda_2(1-z_2)], \tag{18}$$

where $\gamma_1(\cdot)$ is the LST of the busy period of a standard M/G/1 queue with arrival rate λ_1 and LST of the service time $B_1(\cdot)$ (see, e.g., Takács [12]). Thus, $\xi(\cdot)$ can be seen as the PGF of the number of type-2 arrivals during a busy period of queue 1.

By analyticity of $p_1(z_1,z_2)$, the right-hand side of (14) should vanish when $z_1=\xi(z_2)$, i.e.,

$$\alpha(z)=\frac{D(\xi(z),z)+R_1(\xi(z),z)\sum_{j=1}^{k-1}\gamma_j(\xi(z),z)p_{2,k-j}(\xi(z),0)-(R_1(\xi(z),z)-1)\sum_{j=1}^{k-1}\gamma_j(0,z)p_{2,k-j}(0,0)}{z^k-\beta(\xi(z),z)}, \tag{19}$$

and, thus, $\alpha(z)$ is formulated as a function of the unknown functions $p_{2,j}(\cdot,0)$, $j=1,2,\ldots,k-1$.

To eliminate these unknown functions, we differentiate the numerator and denominator of (11) j times with respect to z_2 and, by L'Hospital's rule, we obtain the following recursion, for $j=1,2,\ldots,k-1$,

$$p_{2,j}(x,0)=\sum_{l=1}^{j}\frac{c_l\frac{d^{j-l}}{dy^{j-l}}[S_2^j(x,y)R_2(x,y)]\Big|_{y=0}}{l!(j-l)!}-\sum_{l=1}^{j-1}\frac{\frac{d^l}{dy^l}[S_2^l(x,y)]\Big|_{y=0}}{l!}p_{2,j-l}(x,0), \tag{20}$$

where $c_l=\frac{d^l}{dy^l}[c_0r_2y+\alpha(y)]\Big|_{y=0}$, $l=1,2,\ldots,k-1$.

By (20) we can write $p_{2,j}(\cdot,0)$ as a function of the unknown constants c_j, $j=0,1,\ldots,k-1$. Moreover, with the help of (11), (14) and (19) the PGFs $p_{2,j}(\cdot,\cdot)$, $p_1(\cdot)$ and $\alpha(\cdot)$ can be expressed in terms of these constants as well. The problem of finding these PGFs is, thus, reduced to finding the unknown constants c_j, which can be computed as follows.

Given that (4) holds, the denominator of (19) has exactly k zeros on or within the unit circle (see Adan *et al.* [2]). Since $\alpha(z)$ is bounded in $|z|\le 1$, the zeros in the numerator must be canceled by corresponding zeros in the denominator. One of the zeros equals one and leads to a trivial equation. However, the normalization condition provides an additional equation and, therefore, we have a set of k linear equations. By making the assumption that the k roots of $z^k=\beta(\xi(z),z)$ on or within the unit circle are all distinct, this set of equations has a unique solution for c_j, $j=0,2,\ldots,k-1$. This completes the determination of the PGFs of the queue length distributions at service completion epochs and, hence, in the sequel we assume that these PGFs are known.

Remark 3.1. *It is interesting to note that the function $\beta(\xi(z),z)$ is the PGF of the number of type-2 customers generated in a cycle for queue 2 in which the maximum of k customers is served.* □

3.2 Marginal queue lengths at arbitrary instants

From the results of the previous subsection, we can obtain expressions for the PGF $q_i(\cdot)$ of the marginal queue size distributions of queue i at type-i departure epochs, i.e.,

$$q_1(z)=\frac{p_1(z,1)}{r_1}, \quad \text{and} \quad q_2(z)=\frac{\sum_{j=1}^{k}p_{2,j}(1,z)}{r_2}. \tag{21}$$

By using a standard level crossing argument, in combination with PASTA, it can be shown that the marginal queue length distribution of queue i at type-i departure epochs and at arbitrary instants in time are the same. Hence, the PGFs fore these marginal distributions are given by (21).

Remark 3.2. *From (21) we can easily obtain the LST $W_i(\cdot)$ of the sojourn time distribution of a type-i customer. Since the number of type-i customers left behind by a tagged type-i customer equals the number of customers arrived during the sojourn time of this tagged customer, we have*

$$W_i(z) = q_i\left(1 - \frac{z}{\lambda_i}\right), \qquad i = 1, 2, \tag{22}$$

which is known as the distributional form of Little's law (see, e.g., Keilson and Servi [7]). □

4 Application

In this section, we use the analysis of the present paper to evaluate a stochastic two-item single-capacity production system. After the description of this application in Subsection 4.1, we present a numerical illustration in Subsection 4.2.

4.1 Stochastic economic lot scheduling problem

Consider a system with one single production capacity for two products, in which there is an infinite stock space for each product and raw material is always available. Demands for the two products arrive according to stationary and mutually independent Poisson processes. Demand that cannot be satisfied directly from stock is backlogged until the product becomes available after production. The individual products are produced make-to-stock in a cyclical order with stochastic production times. Stochastic setup times occur *before* the start of the production of a product. For an arbitrary number of products, this setting is often referred to as the *stochastic economic lot scheduling problem* (see Winands *et al.* [13], for a survey).

For product 1, a standard *base-stock* policy is implemented, i.e., when production is commenced, the machine will continue production until a pre-defined target stock level b_1 has been reached. For product 2, a *quantity-limited base-stock policy* is used. That is, when the machine starts production, it will continue production until either the base-stock level b_2 has been reached or a maximum number k of products has been produced. The setup times are assumed to be state-dependent, i.e., no setup for a product is incurred when there is no *shortfall* (no outstanding production orders). When both products have no shortfall, the machine is turned off and is turned on again upon demand arrival. When a type-2 batch of size k has been been produced and product 1 has no shortfall, a new type-2 batch up to k products is started after a new setup time.

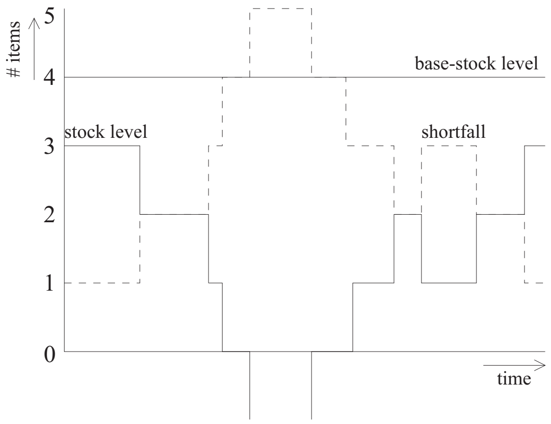

Figure 2: Example of the relation between base-stock level, stock level and shortfall.

Product information		
Parameter	**Product** 1	**Product** 2
Demand	Poisson(0.1)	Poisson(0.5)
Processing times	Exp(1.0)	Exp(1.0)
Setup times	Det(1.0)	Det(0.3)
Holding costs	5	1
Backlogging cost	50	10

Table 1: Detailed product information.

Output					
k	b_1^*	**Marginal costs product** 1	b_2^*	**Marginal costs product** 2	**Total costs**
1	1	6.0	9	9.6	15.7
2	1	6.4	6	6.7	13.1
3	1	6.8	6	6.1	12.9
4	1	7.3	5	5.6	12.9
5	1	7.8	5	5.3	13.1
∞	2	11.2	4	4.0	15.2

Table 2: The marginal costs, the total costs and the optimal base-stock levels as a function of the quantity limit k.

For given values of the base-stocks, b_1 and b_2, and the quantity limit, k, the steady-state stock level distribution I_i for product i is given by (see Figure 2)

$$I_i = b_i - L_i, \qquad i = 1, 2, \tag{23}$$

where L_i denotes the steady-state shortfall of product i. The shortfall of a product is independent of the base-stock levels. It is easily verified that the shortfall distribution of product i is identical to the queue length distribution of queue i in the queueing model of the present paper. Hence, by the procedure presented in Section 3, in combination with (23), the steady-state stock level for both products can be computed. With these distributions, various performance measures of interest can be computed.

In the sequel of the paper, we consider the total expected costs C, i.e., the sum of holding and backlogging costs. For these individual cost components, we assume as simple linear structure. That is, the cost rate function for product i as a function of the stock level x equals

$$c_i(x) = \begin{cases} h_i x, & x > 0, \\ p_i x, & x < 0. \end{cases} \tag{24}$$

With the help of (23), the total expected costs C can be written as follows

$$C = \mathbb{E}[c_1(I_1) + c_2(I_2)] = \mathbb{E}[c_1(b_1 - L_1) + c_2(b_2 - L_2)]. \tag{25}$$

For a given value of k, it can be shown that the optimal base-stock levels b_i^* are given by

$$b_i^* = \min\{n \in \mathbb{N} | P[I_i \leq n] \geq \frac{p_i}{p_i + h_i}\}, \qquad i = 1, 2, \tag{26}$$

which is recognized as the solution of a standard newsboy problem. Finally, we note that we do not consider the optimization of the quantity limit k here.

4.2 Numerical illustration

We now present an example, which illustrates the value of the procedure of Section 3 in the evaluation of the described production system. As described before the derived PGFs have to be finished off with a number of zeros, which are numerically computed by using the Chaudhry QROOT software package [3]. We, then, use the method presented by Abate and Whitt [1] to numerically invert these PGFs. Unfortunately, for large quantity limits, severe numerical problems have been encountered in the procedure. Therefore, we confine ourselves to cases with small limits.

Suppose that product 1 is a slow-moving product with high costs. Product 2 is of secondary importance compared to the first product, but this product has relatively high demand. Table 1 shows the detailed specifications for these two products. Notice that if follows from (4) that the stability condition for this example reads $k > 0.522$.

Table 2 shows the marginal costs, the total costs and the optimal base-stock levels as a function of the quantity limit k. The special case of $k = \infty$ amounts to a two-product model with exhaustive base-stock policy for both products. In this case, we do not use the procedure of Section 3, but we have implemented the standard results for exhaustive polling models (see, e.g., Borst [4]). Several conclusions can be drawn from this table. Firstly, we observe that the marginal costs for product 2 are decreasing in the quantity limit, whereas the marginal costs for product 1 increase with this limit. Secondly, the optimal value of the quantity limit with respect to total costs is approximately equal to 3. For smaller limits the amount of capacity available for production is too low, while larger limits lead to more variable cycle lengths. Finally, the optimal base-stock levels for product 2 are non-increasing in the quantity limit, while the optimal base-stock levels for product 1 are non-decreasing in this limit.

Acknowledgement

The authors would like to thank Onno Boxma for valuable comments.

References

[1] Abate, J., Whitt, W., (1992). *Numerical inversion of probability generating functions* (Operations Research Letters, vol. 12, no. 4, pp. 245-251).

[2] Adan, I.J.B.F., Leeuwaarden, J.S.H. van, Winands, E.M.M., (2005). *On the application of Rouché's theorem in queueing theory* (SPOR-report, Eindhoven University of Technology, Eindhoven).

[3] Chaudhry, M.L., (1994). *QROOT Software Package* (A&A Publications, Kingston).

[4] Borst, S.C., (1996). *Polling Systems* (Centrum voor Wiskunde en Informatica, Amsterdam).

[5] Federgruen, A., Katalan, Z., (1998). *Determining production schedules under base-stock policies in single facility multi-item production systems* (Operations Research, vol. 46, no. 6, pp. 883-898).

[6] Ibe, O.C., Cheng, X., (1988). *Stability conditions for multiqueue systems with cyclic service* (IEEE Transactions Automatic Control, vol. 33, no. 1, pp. 102-103).

[7] Keilson, J., Servi, L.D., (1990). *The distributional form of Little's law and the Fuhrmann-Cooper decomposition* (Operations Research Letters, vol. 9, no. 4, pp. 239-247).

[8] Lee, D.S., (1996). *A two-queue model with exhaustive and limited service disciplines* (Stochastic Models, vol. 12, no. 2, pp. 285-305).

[9] Ozawa, T., (1990). *Alternating service queues with mixed exhaustive and K-limited services* (Performance Evaluation, vol. 11, pp. 165-175).

[10] Ozawa, T., (1997). *Waiting time distribution in a two-queue model with mixed exhaustive and gated-type K-limited services* (Proceedings of International Conference on the Performance and Management of Complex Communication Networks, Tsukuba, pp. 231-250).

[11] Resing, J.A.C., (1993). *Polling systems and multitype branching processes* (Queueing Systems, vol. 13, pp. 409-426).

[12] Takács, L., (1968). *Two queues attended by a single server* (Operations Research, vol. 16, no. 3, pp. 639-650).

[13] Winands, E.M.M, Adan, I.J.B.F., Houtum, G.J. van, (2005). *The stochastic economic lot scheduling problem: a survey* (BETA WP-133, Beta Research School for Operations Management and Logistics, Eindhoven).