

Customer Waiting Times in an (R, S) Inventory System with Compound Poisson Demand

M. C. van der Heijden¹ and A. G. De Kok²

Abstract: Besides service level and mean physical stock, customer waiting time is an important performance characteristic for an inventory system. In this paper we discuss the calculation of this waiting time in case a periodic review control policy with order-up-to-level S is used and customers arrive according to a Poisson process. For the case of Gamma distributed demand per customer, we obtain (approximate) expressions for the waiting time characteristics. The approach clearly differs from the traditional approaches. It can also be used to obtain other performance characteristics such as the mean physical stock and the service level.

1 Introduction

The (R, S) policy is one of the basic single-stockpoint inventory control policies. Under this policy the inventory position, defined as physical stock plus lots on order minus backorders, is reviewed each R^{th} time unit. At a moment of review, a replenishment order is issued such that the inventory position immediately after the order is initiated equals S .

Basic analysis of the (R, S) inventory system can be found in many standard textbooks (see e.g. Hadley and Whitin (1963) and Silver and Peterson (1985)). These analyses mainly focus on the calculation of certain performance measures and the choice of optimal values for R and S (given a certain cost structure) in case backorders occur very infrequently. Performance measures considered are e.g. the mean physical stock and the service level (defined as the fraction of demand satisfied directly from stock on hand). Approximations for these performance characteristics which are broader applicable (for high as well as low service levels), can be found in De Kok (1990).

Less attention is paid to customer waiting times, although waiting time characteristics are interesting as performance measure in single stockpoint models

¹ M. C. van der Heijden, Centre for Quantitative Methods, P.O.Box 414, 5600 AK Eindhoven, The Netherlands.

² A. G. De Kok, Eindhoven University of Technology, Dept. of Mathematics and Computer Science, P.O.Box 513, 5600 MB Eindhoven, The Netherlands.

and as key variable in the analysis of locally controlled multi-echelon inventory systems. Literature on customer waiting times mainly focuses on multi-echelon inventory systems with continuous review. For example, Van Beek (1981) gives a heuristic analysis of the continuous review, fixed order size (b, Q) -model (i.e., a fixed amount Q is ordered each time the inventory position drops below the level b). His analysis is based on the replenishment cycle approach, e.g. the analysis of the inventory behaviour between the arrival of two consecutive replenishment orders. Svoronos and Zipkin (1988) present a more detailed analysis of a similar model.

Results on customer waiting times in periodic review systems are not available as far as we know. For these type of systems, a traditional approach like the replenishment cycle approach cannot be applied straightforwardly. A reason for this is the fact that the analysis of customer waiting times in continuous review systems is based on Little's formula for the relation between the mean backlog and the mean waiting time (see e.g. Svoronos and Zipkin (1988)). In contrast with continuous review systems, for periodic review systems a tractable expression for the mean backlog is not easily obtained using the replenishment cycle approach.

Besides this, Little's formula is only applicable in case of fixed demand per customer. If the demand per customer is stochastic, Little's formula gives a relation between the mean customer waiting time and the mean number of backorders (and *not* the mean backlog). Simply multiplying the mean number of backorders with the mean customer demand may introduce significant errors. As a typical example, in the first of our numerical experiments (see Section 6) we found that the mean number of backorders equals 0.47. Multiplying this with the mean customer demand ($E[D] = 10$) yields that the mean backlog is approximately 4.7. However, the true mean backlog equals 5.2. This error of about 10% is transferred to the mean waiting time as well.

Because of these problems we developed an alternative approach for the (R, S) system with compound Poisson customer demand. We start with a specific initial system state at time 0 and derive expressions for the waiting time characteristics of a customer arriving at time t . By letting $t \rightarrow \infty$ we obtain the stationary waiting time characteristics.

The remaining part of this paper is structured as follows. In the next section we describe the model in more detail and we give the basic notation. Key expressions for the waiting time characteristics are derived in Section 3. These approximate expressions are further elaborated for the special case of Gamma distributed demand per customer in Section 4. To illustrate the power of our approach, we derive expressions for two other performance measures, the mean physical stock and the fraction of demand delivered directly from stock (Section 5). We note that for these performance measures other (approximate) methods are available. In Section 6 we compare our approximations to Monte Carlo simulation results. Finally, conclusions are given in Section 7.

2 Model and Basic Notation

As mentioned in the introduction, we consider the (R, S) inventory system in which the inventory position is reviewed every R time units. At a moment of review an order is released to raise the inventory position to the level S . The lead time of an order has a probability distribution function $G(y)$. Customers arrive at the system according to a Poisson process. The demand per customer is distributed according to the probability distribution function $F(x)$. When a customer arrives and sufficient stock is available to satisfy his demand, the customer is served immediately. Otherwise the available stock is supplied, the remaining demand is backlogged and the customer waits until his demand is *fully* satisfied. There is an infinite waiting room for the waiting customers. The customers are served according to the First Come, First Served (FCFS) discipline.

Note that a customer's order may be delivered in two parts if the available stock is not sufficient. This approach is used quite often in practise (for example in the authors' company), especially since customer's orders are often replenishment orders issued by retailers outside the organization. In this way the customer service level is improved. However, if the objective would be to minimize the average customer waiting time, a different approach seems to be appropriate. Such an approach should be based on an other priority rule than FCFS, taking into account the demand pattern.

For the system as described above we will derive expressions for the following waiting time characteristics:

- (i) The probability that an arbitrary customer has to wait,
- (ii) the mean waiting time,
- (iii) the coefficient of variation of the waiting time,
- (iv) waiting time probabilities.

We will focus on the system behaviour in the stationary situation. That is, the system has been running for such a long time period that it has reached statistical equilibrium. To derive expressions for the waiting time characteristics in the stationary situation, we use the following basic notation:

- R = Length of the period between two consecutive moments of review ($R > 0$).
- S = Order-Up-To-Level ($S > 0$).
- L_i = Lead time of the i^{th} order. The order lead times are identically distributed and have probability distribution function $G(y)$ with mean μ_L and standard deviation σ_L .
- W_t = Waiting time of a customer arriving at time t .
- W = Waiting time of a customer arriving at an arbitrary point in time.
- π_t = Probability that a customer arriving at time t has to wait.
- π = Probability that a customer arriving at an arbitrary point in time has to wait.

D = Demand of an arbitrary customer. All demands are independent and identically distributed and have a general probability distribution function $F(x)$ with mean μ_D and standard deviation σ_D . The probability density of $F(x)$ is denoted by $f(x)$.

$Q[t]$ = The number of orders delivered in the time interval $[0, t]$.

$N[t]$ = The number of customers arriving in an arbitrary time interval of length t . This number is Poisson distributed with mean λt .

$V[t]$ = The cumulative demand in a time interval of length t .

Other, more specific, notation will be defined when used for the first time. Note that both $N[t]$ and $V[t]$ do not depend on the system state at the start of the time interval because of the Poisson arrivals. Further we assume that customer arrival, customer demand and order lead time are independent processes. Finally, we make the following assumption with respect to the lead times:

Assumption: The orders are delivered in the same order as they are released.

So, an order released at time iR is not delivered before any order released at time jR ($j = 1, \dots, i-1$). This is a common assumption (see e.g. the discussion in Hadley and Whitin (1963)) and in fact it is not very restrictive. If the dispatch of a particular replenishment order is delayed considerably, in practise it is *never* crossed, at most it is combined with the next replenishment order. We observe the resulting lead times which we describe by the probability distribution $G(t)$. An implication is that the lead times are not necessarily independent.

3 Derivation of Basic Expressions

As mentioned in the introduction, the key idea of our approach is to start with a specific initial state of the system at time 0 and to derive expressions for the waiting time characteristics of a customer arriving at time $t > 0$. Letting $t \rightarrow \infty$ we obtain the stationary waiting time characteristics.

Now assume that at the initial time $t = 0$ the system is in the following state: The inventory position equals S . No open orders exist and there are no backorders, therefore the physical stock equals S as well. Future orders will be released at time $t = R$, $t = 2R$, etc. We assume that an order with size 0 is released in case no demand occurred between two consecutive moments of review.

Remark: The probability that no demand occurs between two consecutive review periods is strictly positive because of compound Poisson demand. Therefore it

can be shown that the system still re-enter the initial state within finite time. Hence the initial epoch is a regeneration epoch and therefore our results are independent of the initial system state.

We focus on a specific customer arriving at time t , say $mR \leq t < (m+1)R$ for some nonnegative integer m . We will write for convenience $t = mR + T$ with $T \in [0, R)$. We can find an expression for the probability that this customer has to wait by conditioning on the number of orders delivered until time t . Suppose that i orders are delivered until t , i.e. $Q[t] = i$. Then the customer has to wait if the i orders are not sufficient to cover all demand up to time t including the customer arrived at t . Because the i -th order increased the inventory position at time iR to the level S , this is true if the total demand in the period (iR, t) plus the demand of the customer arriving at t exceeds S . Noting that at time $t = mR + t$ exactly m orders are released, we have $Q[t] \leq m$. Therefore we can write

$$\pi_{T+mR} = \sum_{i=0}^m \Pr\{Q[T+mR] = i\} * \Pr\{V[T+(m-i)R] + D > S\}, \quad 0 \leq T < R \quad (1)$$

Here D denotes the demand of the customer that arrived at time t . Similarly we obtain an expression for the probability that the customer has to wait longer than some $w \geq 0$ by conditioning on the number of orders delivered until time $t + w$. Suppose that i orders are delivered until $t + w$, i.e. $Q[t+w] = i$. Then the waiting time of the customer exceeds w if the i orders are not sufficient to cover all demand up to time t including the customer arrived at t . Because the $(m+1)^{\text{th}}$ order will cover all demand up to time $(m+1)R$ and $t < (m+1)R$, we have

$$\Pr\{W_{T+mR} > w\} = \sum_{i=0}^m \Pr\{Q[T+mR+w] = i\} * \Pr\{V[T+(m-i)R] + D > S\},$$

$$0 \leq T < R \quad \text{and} \quad w \geq 0 \quad (2)$$

To elaborate the expressions (1) and (2) we need the probability distribution of $Q[t]$. Under the assumption that orders do not cross (see Section 2), we have

$$\Pr\{Q[t] \geq i\} = \Pr\{iR + L_i \leq t\} = G(t - iR), \quad i \geq 1 \quad (3)$$

Of course, for $i = 0$ we have $\Pr\{Q[t] \geq 0\} = 1$. Now it is clear that

$$\begin{aligned} \Pr\{Q[t] = i\} &= \Pr\{Q[t] \geq i\} - \Pr\{Q[t] \geq i + 1\} \\ &= G(t - iR) - G(t - (i + 1)R) , \quad i \geq 1 \end{aligned} \quad (4a)$$

$$\Pr\{Q[t] = 0\} = 1 - G(t - R) \quad (4b)$$

with the convention that $G(x) = 0$ for $x < 0$. Substitution of (4a) and (4b) in (1) and (2) while changing the order of summation ($i := m - i$) yields

$$\begin{aligned} \pi_{T+mR} &= \sum_{i=0}^{m-1} [G(T + iR) - G(T + (i - 1)R)] * \Pr\{V[T + iR] + D > S\} \\ &\quad + [1 - G(T + (m - 1)R)] * \Pr\{V[T + mR] + D > S\} , \quad 0 \leq T < R \end{aligned} \quad (5)$$

and

$$\begin{aligned} \Pr\{W_{T+mR} > w\} &= \sum_{i=0}^{m-1} [G(T + iR + w) - G(T + (i - 1)R + w)] \\ &\quad * \Pr\{V[T + iR] + D > S\} + [1 - G(T + (m - 1)R + w)] \\ &\quad * \Pr\{V[T + mR] + D > S\} , \quad 0 \leq T < R \quad \text{and} \quad w \geq 0 . \end{aligned} \quad (6)$$

We can derive an expression for the stationary waiting time distribution from (5) and (6). It is easy to show that for compound Poisson demand the arrival epoch T of an arbitrary customer arriving between time mR and time $(m + 1)R$ is uniformly distributed over the interval $[mR, (m + 1)R)$. Defining π_m and W_m as the waiting probability respectively the waiting time of an arbitrary customer arriving between the m^{th} and the $(m + 1)^{\text{th}}$ moment of review, we have

$$\pi_m = R^{-1} * \int_0^R \pi_{T+mR} dT \quad (7)$$

and

$$\Pr\{W_m > w\} = R^{-1} * \int_0^R \Pr\{W_{T+mR} > w\} dT . \quad (8)$$

Now we derive an expression for the waiting probability π . Combining (5) and (7) and letting $m \rightarrow \infty$ we obtain after elementary algebra

$$\pi = R^{-1} \int_0^\infty [G(t) - G(t-R)] * \Pr\{V[t] + D > S\} dt . \tag{9}$$

Substituting $\Pr\{V[t] + D > S\} = 1 - \Pr\{V[t] + D \leq S\}$ and taking into account that $G(t) = 0$ for $t < 0$, we have

$$\pi = 1 - R^{-1} * \left\{ \int_0^\infty [1 - G(t)] * \Pr\{V[t+R] + D \leq S\} dt + \int_0^R \Pr\{V[t] + D \leq S\} dt - \int_0^\infty [1 - G(t)] * \Pr\{V[t] + D \leq S\} dt \right\} . \tag{10}$$

Similar algebra using (6) and (8) gives an expression for $\Pr\{W > w\}$:

$$\Pr\{W > w\} = R^{-1} * \left\{ M_2(w) + \int_{M_1(w)}^w [1 - G(t)] dt - \int_{M_1(w)}^\infty [1 - G(t)] * \Pr\{V[t+R-w] + D \leq S\} dt - \int_0^{M_2(w)} \Pr\{V[t] + D \leq S\} dt + \int_w^\infty [1 - G(t)] * \Pr\{V[t-w] + D \leq S\} dt \right\} , \tag{11}$$

where $M_1(w) = \text{Max}\{w - R, 0\}$ and $M_2(w) = \text{Max}\{R - w, 0\}$.

To obtain the mean and coefficient of variation of the waiting time distribution, we note that

$$E[W^k] = \int_0^\infty k w^{k-1} * \Pr\{W > w\} dw .$$

After some algebra we obtain the following expressions for the first two moments of the customer waiting time:

$$E[W] = R/2 + \mu_L - R^{-1} * \int_0^R (R-t) * \Pr\{V[t] + D \leq S\} dt - R^{-1} * \int_0^\infty [1 - G(w)] * \int_w^{w+R} \Pr\{V[t] + D \leq S\} dt dw \tag{12}$$

and

$$\begin{aligned}
 E[W^2] = & R^2/3 + E[L^2] + R * E[L] - R^{-1} * \int_0^R (R-t)^2 * \Pr\{V[t] + D \leq S\} dt \\
 & - 2 * \int_0^\infty [1 - G(w)] * \int_0^{w+R} \Pr\{V[t] + D \leq S\} dt dw \\
 & + 2 * R^{-1} \int_0^\infty [1 - G(w)] * \int_w^{w+R} (t-w) * \Pr\{V[t] + D \leq S\} dt dw . \quad (13)
 \end{aligned}$$

To evaluate the basic expressions (10) – (13), we need $\Pr\{V[t] + D \leq S\}$ as a function of t . In general it is not easy to get computationally tractable expressions for this function. However, in case the demand per customer is Gamma distributed we end up with numerically attractive expressions for π , $E[W]$ and $E[W^2]$. This is discussed in the next section.

4 Expressions in Case of Gamma Distributed Demand per Customer

In this section we discuss the basic expressions for the case of Gamma distributed demand per customer. We focus on the waiting probability π and the first two moments of the waiting time distribution $E[W]$ and $E[W^2]$ respectively. Waiting time probabilities can be approximated by fitting an appropriate probability distribution function to the first two moments of the conditional waiting time (i.e. the waiting time given that it is nonzero). For example, fitting a mixture of Erlang distributions has shown to be very useful for practical applications (see e.g. Tijms (1986)). However, it is possible to calculate waiting time probabilities using the approach discussed in the previous section (see Appendix 2).

Let us focus on tractable expressions for π , $E[W]$ and $E[W^2]$. First we need an expression for $\Pr\{V[t] + D \leq S\}$. Because the total demand in a period of length t is a compound Poisson process, we can easily write down the probability that $V[t] + D$ does not exceed x for some $x \geq 0$:

$$\Pr\{V[t] + D \leq x\} = \sum_{i=0}^\infty (\lambda t)^i e^{-\lambda t} / i! * F^{(i+1)}(x) , \quad x \geq 0 , \quad (14)$$

where $F^{(i)}(x)$ denotes the i -fold convolution of $F(x)$. Substitution of (14) in (10) yields the following expression for the waiting probability:

$$\pi = 1 - R^{-1} * \sum_{i=0}^{\infty} F^{(i+1)}(S) * \left\{ \left[1 - \sum_{j=0}^i (\lambda R)^j e^{-\lambda R} / j! \right] / \lambda - Z_i(\lambda) + \sum_{j=0}^i (\lambda R)^j e^{-\lambda R} / j! * Z_{i-j}(\lambda) \right\}$$

where

$$Z_i(\lambda) = \int_0^{\infty} [1 - G(t)] * (\lambda t)^i e^{-\lambda t} / i! dt . \tag{15}$$

For the first two moments of the waiting time distribution we obtain respectively

$$E[W] = R/2 + E[L] - (R\lambda)^{-1} * \sum_{i=0}^{\infty} F^{(i+1)}(S) * \left\{ R * \left[1 - \sum_{j=0}^i (\lambda R)^j e^{-\lambda R} / j! \right] - (i+1) * \left[1 - \sum_{j=0}^{i+1} (\lambda R)^j e^{-\lambda R} / j! \right] / \lambda + \sum_{j=0}^i \left[1 - \sum_{k=0}^{i-j} (\lambda R)^k e^{-\lambda R} / k! \right] * Z_j(\lambda) \right\} \tag{16}$$

and

$$E[W^2] = R^2/3 + E[L^2] + R * E[L] - \lambda^{-1} * \sum_{i=0}^{\infty} F^{(i+1)}(S) * \left\{ R * \left[1 - \sum_{j=0}^i (\lambda R)^j e^{-\lambda R} / j! \right] - 2(i+1) * \left[1 - \sum_{j=0}^{i+1} (\lambda R)^j e^{-\lambda R} / j! \right] / \lambda + (i+1)(i+2) * \left[1 - \sum_{j=0}^{i+2} (\lambda R)^j e^{-\lambda R} / j! \right] / (R\lambda^2) + 2E[L] - 2 * \sum_{j=0}^i \left[\sum_{k=0}^{i-j} (\lambda R)^k e^{-\lambda R} / k! \right] * Z_j(\lambda) - 2 * \sum_{j=0}^{i+1} (i+1-j) * \left[1 - \sum_{k=0}^{i+1-j} (\lambda R)^k e^{-\lambda R} / k! \right] * Z_j(\lambda) / \lambda R \right\} . \tag{17}$$

Now assume that $F(x)$ is a Gamma distribution with parameters θ and r :

$$F(x) = \int_0^x \theta^r y^{r-1} e^{-\theta y} / \Gamma(r) dy .$$

Then the convolutions $F^{(i)}(x)$ are Gamma distributions with parameters θ and $i * r$ and the values $F^{(i+1)}(S)$ can be calculated numerically using available routines (cf. Press, Flannery, Teukolsky and Vetterling (1986)). The waiting time characteristics can be calculated by truncating the infinite sum after some integer N_ε when the relative change in π (respectively $E[W]$ and $E[W^2]$) is smaller than some predefined value ε . The only part remaining is the calculation of the values $Z_i(\lambda)$ for $i = 1, 2, \dots, N_\varepsilon$. In principle these values can be calculated using standard numerical integration routines for a wide variety of probability distribution functions for the lead time $G(t)$. Especially, Laguerre integration is an efficient numerical method for this type of integrals (see e.g. Van Hoorn (1984)). However, for some specific distribution functions analytical expressions for $Z_i(\lambda)$ can be obtained (see Appendix 1 for deterministic, Erlang distributed and Hyperexponential distributed lead times).

5 Other Performance Measures

The approach presented in this paper is suitable for deriving approximations of other relevant performance characteristics. As an example we derive expressions for the mean physical stock and the fraction of the demand delivered directly from stock (denoted by β). We remark that other methods to approximate these performance measures are available (see De Kok (1990)).

5.1 The Mean Physical Stock

First we derive an expression for the probability that the physical stock at time $t = T + mR$ exceeds x . We condition on $Q[t]$, the number of orders arrived until t . Given that $Q[t] = i$, we have that the physical stock at time t exceeds x if the cumulative demand in the period $[iR, t]$ does not exceed $S - x$. So,

$$\Pr\{I_{T+mR} \geq x\} = \sum_{i=0}^m \Pr\{Q[T+mR] = i\} * \Pr\{V[T+(m-i)R] \leq S-x\} ,$$

$$0 < x \leq S ,$$

where I_{T+mR} denotes the physical stock at time $T+mR$. Analogous arguments as the ones used in Section 3 (integrating over T and letting $m \rightarrow \infty$) yields

$$\Pr \{I \geq x\} = R^{-1} * \int_0^\infty [G(t) - G(t-R)] * \Pr \{V[t] \leq S-x\} dt, \quad 0 < x \leq S. \quad (18)$$

The mean physical stock can be derived by integration over x :

$$E[I] = R^{-1} * \int_0^\infty \int_0^S [G(t) - G(t-R)] * \Pr \{V[t] \leq S-x\} dx dt. \quad (19)$$

5.2 The Fraction of Demand Satisfied Directly from Stock on Hand

To obtain this service level, we observe that

$$\beta = E[D_0] / \mu_D, \quad (20)$$

where D_0 denotes the demand of an arbitrary customer which is delivered directly from stock. D_0 exceeds x if both the customer demand D and the available physical stock exceed x . So we have

$$\Pr \{D_0 > x\} = [1 - F(x)] * \Pr \{I \geq x\}. \quad (21)$$

Combining (18), (20) and (21) yields

$$\beta = (R\mu_D)^{-1} * \int_0^\infty [G(t) - G(t-R)] * \int_0^S [1 - F(x)] * \Pr \{V[t] \leq S-x\} dx dt. \quad (22)$$

Remark that the expressions for the mean physical stock and the service level can be computed explicitly in case of Gamma distributed demand per customer.

6 Numerical Results

In this section we examine the accuracy of the approximating expressions by comparison to discrete event simulation. We consider the following parameter

values. As normalizations we take $R = 1$ and $\mu_D = 10$. The arrival rate of customers equals $\lambda = 5$ or 25 per time unit. The demand per customer is gamma distributed with squared coefficient of variation $c_D^2 = 1.5$. We choose the mean lead time of an order as $\mu_L = 0.5$ and 2. Both deterministic lead times ($c_L^2 = 0$) and Erlang distributed lead times ($c_L^2 = 0.25$) are considered. Note that for deterministic lead times the expressions are exact, because then orders do not cross. We did not consider hyperexponential distributed lead times, because lead time variation is not very large in practise. We choose the order-up-to level S such, that the service level β approximately equals 0.75, 0.90 and 0.99 (where S is rounded to an integer value). This is done using a standard numerical search procedure and the approximate method as described in the previous sections. The low service level ($\beta = 0.75$) is considered because one usually finds in practice that at intermediate stockpoints in a logistic network target service levels are in the range of 0.60 to 0.80.

For all parameter combinations we calculated the waiting probability, the mean and squared coefficient of variation of the conditional waiting time (i.e. the waiting time given that it is strictly positive), the actual service level β and the mean physical stock. We compared this to results from discrete event simulation. For each case we simulated 1000000 customers. The Tables 1 – 3 show the results for the service levels 0.75, 0.90 and 0.99 respectively.

The tables show that for most cases the (approximate) values are very close to the simulation results. Significant deviations mainly occur for relatively long, stochastic lead times. A probable cause is the assumption that orders cannot overtake each other.

Table 1. Results for service level $\beta = 0.75$ ($R = 1$, $\mu_D = 10$ and $c_D^2 = 1.5$)

λ	μ_L	c_L^2	S		π_w	$E[W W>0]$	$c_{W W>0}^2$	Actual β	Mean Stock
5	0.5	0	85	Sim	0.233	0.41	0.55	0.751	40
				Appr	0.231	0.41	0.55	0.751	40
5	0.5	0.25	86	Sim	0.232	0.45	0.60	0.753	42
				Appr	0.231	0.45	0.60	0.753	42
5	2.0	0	173	Sim	0.239	0.63	0.58	0.750	56
				Appr	0.238	0.63	0.58	0.750	56
5	2.0	0.25	181	Sim	0.240	1.00	0.78	0.752	68
				Appr	0.241	1.00	0.75	0.750	69
25	0.5	0	332	Sim	0.246	0.25	0.59	0.748	97
				Appr	0.244	0.25	0.59	0.750	98
25	0.5	0.25	340	Sim	0.246	0.31	0.69	0.748	109
				Appr	0.245	0.31	0.69	0.750	110
25	2.0	0	731	Sim	0.248	0.33	0.63	0.747	126
				Appr	0.246	0.34	0.63	0.749	127
25	2.0	0.25	802	Sim	0.249	0.89	0.91	0.750	230
				Appr	0.247	0.87	0.82	0.750	231

Table 2. Results for service level $\beta = 0.90$ ($R = 1, \mu_D = 10$ and $c_D^2 = 1.5$)

λ	μ_L	c_L^2	S		π_w	$E[W W>0]$	$c_{W W>0}^2$	Actual β	Mean Stock
5	0.5	0	139	Sim	0.046	0.31	0.67	0.950	90
				Appr	0.046	0.31	0.66	0.949	90
5	0.5	0.25	143	Sim	0.046	0.35	0.73	0.950	94
				Appr	0.046	0.35	0.71	0.950	94
5	2.0	0	245	Sim	0.047	0.46	0.70	0.950	121
				Appr	0.047	0.46	0.70	0.949	121
5	2.0	0.25	283	Sim	0.047	0.85	0.95	0.951	160
				Appr	0.048	0.83	0.86	0.950	160
25	0.5	0	455	Sim	0.047	0.17	0.72	0.950	207
				Appr	0.048	0.17	0.71	0.950	207
25	0.5	0.25	487	Sim	0.048	0.23	0.85	0.950	239
				Appr	0.048	0.24	0.82	0.950	240
25	2.0	0	889	Sim	0.048	0.23	0.76	0.950	265
				Appr	0.049	0.24	0.74	0.950	267
25	2.0	0.25	1173	Sim	0.052	0.81	1.17	0.948	555
				Appr	0.049	0.74	0.90	0.950	557

Table 3. Results for service level $\beta = 0.99$ ($R = 1, \mu_D = 10$ and $c_D^2 = 1.5$)

λ	μ_L	c_L^2	S		π_w	$E[W W>0]$	$c_{W W>0}^2$	Actual β	Mean Stock
5	0.5	0	186	Sim	0.009	0.26	0.72	0.990	136
				Appr	0.009	0.26	0.72	0.990	136
5	0.5	0.25	193	Sim	0.009	0.30	0.78	0.990	143
				Appr	0.009	0.30	0.78	0.990	143
5	2.0	0	304	Sim	0.009	0.39	0.75	0.990	179
				Appr	0.009	0.39	0.76	0.990	179
5	2.0	0.25	371	Sim	0.010	0.81	1.13	0.990	246
				Appr	0.009	0.75	0.90	0.990	246
25	0.5	0	547	Sim	0.009	0.14	0.77	0.991	297
				Appr	0.009	0.14	0.77	0.990	297
25	0.5	0.25	603	Sim	0.009	0.21	0.95	0.990	353
				Appr	0.010	0.20	0.88	0.990	354
25	2.0	0	1007	Sim	0.009	0.19	0.80	0.990	381
				Appr	0.010	0.19	0.80	0.990	382
25	2.0	0.25	1491	Sim	0.012	0.84	1.74	0.988	865
				Appr	0.010	0.68	0.94	0.990	868

We found that the computer time required for our analytical method stays within reasonable bounds, although a lot of gamma integrals have to be evaluated. As

a test we calculated the waiting probability and the first two moments of the waiting time for 112 cases. The average runtime on a Personal Computer with 80386 microprocessor was about 2 seconds. The maximal runtime was about 10 seconds. The runtime required increases as the ratio S/μ_D increases, because then a lot of terms of the infinite sums in the expressions (15)–(17) have to be evaluated. So, the runtime gets high in two cases: Firstly in case of a very high service level (when waiting times are irrelevant). Secondly, when the lead time of an order is very long compared to the interarrival times of customers. In case the runtime gets too high, application of the PDF-method as suggested by De Kok (1990) can be considered.

7 Conclusions

In this paper we presented a method to calculate waiting time characteristics in an (R, S) inventory system with compound Poisson demand. For the special case of Gamma distributed demand of customers we can compute explicitly the performance measures. We compared the approximate expressions to Monte Carlo simulation results. The approximations appear to be very accurate.

We expect that the approach presented in this paper can be applied to related models and other performance characteristics.

Appendix 1

Calculating $Z_i(\lambda)$ for some special cases.

The function $Z_i(\lambda)$ as defined at formula (15) can be calculated easily for some specific lead time distributions. In this appendix we will discuss three important cases: Deterministic lead times, lead times having a mixture of Erlang distributions and Hyperexponential distributed lead times.

A. Deterministic Lead Times

When the lead time of an order is deterministic, we have that $1 - G(w) = 1$ for $w \leq \mu_L$ and 0 for $w > \mu_L$. It is easy to see that for this case

$$Z_i(\lambda) = \left[1 - \sum_{j=0}^i (\lambda \mu_L)^j e^{-\lambda \mu_L / j!} \right] / \lambda .$$

B. Erlang Lead Times

Consider the case where the lead time is distributed according to a mixture of Erlang distributions with the same scale parameter α , defined by

$$1 - G(t) = q * \sum_{j=0}^{n-2} (\alpha t)^j e^{-\alpha t} / j! + (1 - q) * \sum_{j=0}^{n-1} (\alpha t)^j e^{-\alpha t} / j! , \quad t \geq 0 .$$

This versatile density is often used in practical applications. It has the property that its squared coefficient of variation is within the range (0, 1], where the coefficient of variation c_x of a positive random variable X is defined as the ratio of its standard deviation σ_x and its mean μ_x :

$$c_x = \sigma_x / \mu_x .$$

Given the mean μ_L and squared coefficient of variation c_L^2 of the lead time, the parameters n , α and q can be solved (see Tijms (1986)). Then we obtain the following expression for $Z_i(\lambda)$:

$$Z_i(\lambda) = \sum_{j=0}^{n-2} \binom{i+j}{i} * \frac{\lambda^i \alpha^j}{(\lambda + \alpha)^{i+j+1}} + (1 - q) * \binom{i+n-1}{i} * \frac{\lambda^i \alpha^{n-1}}{(\lambda + \alpha)^{i+n}} .$$

C. Hyperexponential Lead Times

Finally we consider lead times having a Hyperexponential distribution, defined by

$$1 - G(t) = q * e^{-\alpha_1 t} + (1 - q) * e^{-\alpha_2 t} , \quad t \geq 0 .$$

The Hyperexponential distribution is applicable for lead times having a large coefficient of variation. It has the property that its squared coefficient of variation is larger than 0.5. We obtain the following expression for $Z_i(\lambda)$:

$$Z_i(\lambda) = q * \frac{\lambda^i}{(\lambda + \alpha_1)^{i+1}} + (1 - q) * \frac{\lambda^i}{(\lambda + \alpha_2)^{i+1}} .$$

Appendix 2

Waiting time probabilities.

As mentioned in Section 4, it is possible to calculate waiting time probabilities in case the demand per customer is Gamma distributed. An expression for $\Pr\{W > w\}$ is obtained by substitution of (14) in (11):

$$\begin{aligned} \Pr\{W > w\} = & M_2(w)/R + R^{-1} * \int_{M_1(w)}^w [1 - G(t)] dt - R^{-1} * \sum_{i=0}^{\infty} F^{(i+1)}(S) \\ & * \left\{ \left[1 - \sum_{j=0}^i (\lambda M_2(w))^j e^{-\lambda M_2(w)} / j! \right] / \lambda \right. \\ & + \sum_{j=0}^i (\lambda(R-w))^{i-j} e^{-\lambda(R-w)} / (i-j)! * Y_j(M_1(w); \lambda) \\ & \left. - \sum_{j=0}^i (-\lambda w)^{i-j} e^{\lambda w} / (i-j)! * Y_j(w; \lambda) \right\} \end{aligned}$$

where

$$M_1(w) = \text{Max}\{w - R, 0\} ,$$

$$M_2(w) = \text{Max}\{R - w, 0\}$$

and

$$Y_j(w; \lambda) = \int_w^{\infty} [1 - G(t)] * (\lambda t)^j e^{-\lambda t} / j! dt .$$

To calculate $\Pr\{W > w\}$, we need the values $Y_j(w; \lambda)$ for $j = 1, 2, \dots$ and the integral

$$\int_{M_1(w)}^w [1 - G(t)] dt .$$

In this appendix we deal with the special cases of deterministic lead times, lead times having a mixture of Erlang distributions and hyperexponential distributed lead times.

A. Deterministic Lead Times

For this case we have

$$Y_i(w; \lambda) = \lambda^{-1} * \sum_{j=0}^i \{(\lambda w)^j e^{-\lambda w} / j! - (\lambda \mu_L)^j e^{-\lambda \mu_L} / j!\} , \quad 0 \leq w \leq \mu_L$$

and $Y_i(w; \lambda) = 0$ otherwise. Further we have that

$$\int_{M_1(w)}^w [1 - G(t)] dt = \text{Min}\{w, \mu_L\} - \text{Min}\{M_1(w), \mu_L\} .$$

B. Erlang Lead Times

In case the lead time distribution is a mixture of Erlang distributions, we can show that

$$Y_i(w; \lambda) = \sum_{j=0}^{n-2} \binom{i+j}{i} * \frac{\lambda^i \alpha^j}{(\lambda + \alpha)^{i+j+1}} * \sum_{k=0}^{i+j} [(\lambda + \alpha) w]^k e^{-(\lambda + \alpha) w} / k! \\ + (1 - q) * \binom{i+n-1}{i} * \frac{\lambda^i \alpha^{n-1}}{(\lambda + \alpha)^{i+n}} * \sum_{k=0}^{i+n-1} [(\lambda + \alpha) w]^k e^{-(\lambda + \alpha) w} / k!$$

and

$$\int_{M_1(w)}^w [1 - G(t)] dt = \sum_{i=0}^{n-2} (n - i - 1) * \{[\alpha M_1(w)]^i e^{-\alpha M_1(w)} / i! - (\alpha w)^i e^{-\alpha w} / i!\} / \alpha \\ + (1 - q) * \sum_{i=0}^{n-1} \{[\alpha M_1(w)]^i e^{-\alpha M_1(w)} / i! - (\alpha w)^i e^{-\alpha w} / i!\} / \alpha .$$

C. Hyperexponential Lead Times

For Hyperexponentially distributed lead times we can show that

$$Y_i(w; \lambda) = \frac{q * \lambda^i}{(\lambda + \alpha_1)^{i+1}} * \sum_{j=0}^i \frac{[(\lambda + \alpha_1)w]^j e^{-(\lambda + \alpha_1)w}}{j!} \\ + \frac{(1-q) * \lambda^i}{(\lambda + \alpha_2)^{i+1}} * \sum_{j=0}^i \frac{[(\lambda + \alpha_2)w]^j e^{-(\lambda + \alpha_2)w}}{j!}$$

and

$$\int_{M_1(w)}^w [1 - G(t)] dt = q * [e^{-\alpha_1 M_1(w)} - e^{-\alpha_1 w}] / \alpha_1 + (1-q) * [e^{-\alpha_2 M_1(w)} - e^{-\alpha_2 w}] / \alpha_2.$$

Acknowledgement: We thank two anonymous referees for their constructive comments on our paper.

References

1. De Kok AG (1990) Basics of Inventory Management Models, Research Reports FEW 520 – 525, Tilbury University
2. Hadley G, Whitin TM (1963) Analysis of Inventory Systems. Prentice-Hall, Englewood Cliffs
3. Press WH, Flannery BP, Teukolsky SA, Vetterling WT (1986) Numerical Recipes – the Art of Scientific Computing. Cambridge University Press, New York
4. Silver EA, Peterson R (1985) Decision Systems for Inventory Management and Production Planning. Wiley, New York
5. Svoronos A, Zipkin P (1988) Estimating the Performance of Multi-Level Inventory Systems. Operat Res 36:57 – 72
6. Tijms HC (1986) Stochastic Modelling and Analysis – A Computational Approach. Wiley, Chichester
7. Van Beek P (1981) Modelling and Analysis of a Multi-Echelon Inventory System. Eur J Oper Res 6:380 – 385
8. Van Hoorn MH (1984) Algorithms and Approximations for Queuing Systems. CWI Tract No 8, Amsterdam

Received: August 1990

Revised version received: April 1991