

Multi-item inventory control with full truckloads : a comparison of aggregate and individual order triggering

Citation for published version (APA):

Kiesmuller, G. P. (2008). *Multi-item inventory control with full truckloads : a comparison of aggregate and individual order triggering*. (BETA publicatie : working papers; Vol. 245). Technische Universiteit Eindhoven.

Document status and date:

Published: 01/01/2008

Document Version:

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Multi-item inventory control with full truckloads

A comparison of aggregate and individual order triggering

GP Kiesmüller

Faculty of Technology Management,

Technische Universiteit Eindhoven,

P.O. Box 513,

5600 MB Eindhoven,

The Netherlands

This paper should not be quoted or referred to without the prior written permission of the authors

Multi-item inventory control with full truckloads

A comparison of aggregate and individual order triggering

G. P. Kiesmüller

Faculty of Technology Management,

Technische Universiteit Eindhoven,

P.O. Box 513,

5600 MB Eindhoven, The Netherlands

e-mail: g.p.kiesmueller@tm.tue.nl

Abstract

In this paper we consider the stochastic joint-replenishment problem in an environment where transportation costs are dominant and full truckloads or full container loads are required. One replenishment policy, taking into account capacity restrictions of the total order volume, is the so-called QS policy, where replenishment orders are placed to raise the individual inventory positions of all items to their order-up-to levels, whenever the aggregate inventory position drops below the reorder level. We first provide a method to compute the policy parameters of an QS policy such that item target service levels can be met, under the assumption that demand can be modeled as a compound renewal process. The approximate formulas are based on renewal theoretic results and are tested in a simulation study, revealing a good performance. Second, we compare the QS policy with a simple allocation policy, where replenishment orders are triggered by the individual inventory positions of the items. At the moment when an individual inventory position drops below its item reorder level a replenishment order is triggered and the total vehicle capacity is allocated among all items such that the expected elapsed time before the next replenishment order is maximized. In an extensive simulation study it is illustrated that the QS policy outperforms this allocation policy, standing for lower inventory levels to obtain the same service level. While for identical items the difference between the performance of both policies is negligible, differences can be large for different item characteristics.

keywords: inventory, stochastic modelling, coordinated replenishments

1 Introduction

Many of the stochastic inventory control models which have been studied in the literature are single item models or assume that multiple items are ordered independently of each other, resulting in a large number of small orders. But in real-life situations coordination of orders and joint replenishments often take place in order to achieve economies of scale. Joint replenishment policies are cost effective when there is a fixed major cost for each order and only a minor ordering cost for each item included in the order. However, the body of literature studying the stochastic joint replenishment problem (see for an overview Aksoy and Erenguc (1987) or Goyal and Satir (1989)) is much less than for independent replenishments.

One reason for this may be the fact that the optimal coordinated control policy for the stochastic case is in general unknown and is expected to have a difficult structure (see Ignall (1969)). Therefore, reasonable policies are studied and formulas and algorithms are provided to determine optimal and near-optimal policy parameters. We can distinguish, as in the single item case, between continuous and periodic review policies.

Probably the best known continuous joint replenishment policy is the can-order-policy, defined by three different parameters (S, c, s) for each item, the order-up-to level S , the can-order level c and the must order level s . Whenever the inventory position of an individual item is dropping to or below the must-order level a replenishment order is triggered to raise the item inventory position up to the corresponding order-up-to level S . All items with an inventory position below the can-order level are also included in the replenishment order and their inventory position is also raised to their order-up-to level. This policy is first proposed by Balintfy (1964) and the computation of near optimal parameters is discussed in a series of papers for different demand and leadtime assumptions (see Silver (1974), Silver (1981), Thompson and Silver (1975) and Federgruen et al. (1984)). All the methods use the idea to decompose the N item problem in N independent single item problems by assuming that discounted replenishment opportunities for an item (replenishment moments which are triggered by another item) follow a Poisson process. However, van Eijs (1994) has shown that especially for large major ordering costs the proposed procedures can perform quite bad and do not provide the cost optimal policy parameters.

Improved computational procedures are presented in Melchior (2002) and Schultz and Johansen (1999).

Another sub-class of policies reviews the inventory level periodically. How to compute policy parameters for order-up-to policies when demand follows a Poisson process can be found in Atkins and Iyogun (1988) and for compound Poisson demand in Fung and Lau (2001). A periodic (s, S) policy is discussed in Viswanathan (1997) and a periodic can-order policy in Johanson and Melchior (2003).

All ordering policies mentioned above have in common that the total order volume is variable and full truckloads or containersizes cannot be guaranteed. However, transportation managers of companies owning a private fleet, aim to get a high utilization of their equipment and try to create full truckloads. Further, carrier freight rates are often a function of the order volume and often the cheapest prices are offered for full truckloads (see also Swenseth and Godfrey (2002) for a discussion of freight rates in practice). Moreover, shippers sometimes have so-called transportation capacity reservation contracts with logistics service providers where a fixed transportation capacity is reserved for a guaranteed price. Then ordering policies have to take into account the capacity constraint on the total order volume, but up to now stochastic inventory models including this capacity constraint are rare and results are also very limited.

In this paper we study and compare two inventory control policies where the total order volume is fixed, in the following always mentioned as policies where goods are shipped as full truckloads. We first show how to compute the policy parameters for the so-called QS policy (Renberg and Planche (1967)), which works as follows. The aggregate inventory position of all items is continuously reviewed and when it is equal to the reorder level, orders are triggered for each item to raise all individual inventory positions up to their order-up-to level. In case the difference between the sum of all order-up-to levels minus the reorder level is equal to the capacity of a vehicle, full truckloads are guaranteed. Under the assumption that demand follows a Poisson process Pantumsinchai (1992) shows how to compute the policy parameters such that total costs, composed of major and minor ordering costs, shortage and penalty costs for backorders and inventory holding costs, are minimized. Further, it is shown by means of numerical examples that the policy

outperforms other policies in case of high major ordering costs and low shortage costs, but it has a tendency to incur high shortages since an item can run out of stock while the aggregate inventory position is still above the reorder level. This conclusion is in line with the conclusions of Viswanathan (1997) who additionally states that the QS-policy performs well when items have identical cost and demand parameters.

In contrast to the paper of Pantumsinchai (1992) we model the demand process as a compound renewal process, which enables us to model demand sizes as well as interarrival times of orders with random variables. Moreover, the squared coefficient of variation of these random variables does not have to be equal to one. We provide approximate formulas, based on renewal theoretic results, in order to determine the policy parameters and we test the performance of the approximations. Further, instead of using a cost approach as in Pantumsinchai (1992) and in most of the literature in this field, we use a service level approach, which means that we assume that for each item a target service level, in this case the fill rate, has to be met. This should prevent the tendency towards high shortages for individual items, but then the question arises if high inventory levels occur as a side effect.

Therefore, we compare in the second part of the paper the QS policy with a policy where replenishment orders are triggered based on the individual inventory position of an item. At the moment when an arbitrary item inventory position drops below the item reorder level, a replenishment order is triggered, and the total order volume is allocated to all items in order to guarantee a full truckload. The allocation decision is based on the idea to maximize the time until the next order is triggered, similar to Miltenburg (1985), who studies allocation under the assumption that inventory positions can be modeled as a Wiener Process. We compare this allocation policy with the QS policy in a numerical simulation study and we illustrate that less inventory is needed under a QS policy to obtain the same service level, compared to the allocation policy. The performance difference is dependent on the item characteristics and differences are negligible for identical items.

The structure of the paper is as follows. In section 2 we provide formulas in order to compute the policy parameters of an QS policy under compound renewal demand and we present the results of a simulation study testing the performance of the approximations.

In section 3 we describe a replenishment policy where replenishment orders are triggered by an individual inventory position and we compare this policy with the QS policy. Then, in section 4 we summarize our results and draw conclusions.

2 Aggregate order triggering

2.1 The QS policy

In this paper we assume that full truckloads are always profitable and in order to profit from these economies of scale, inventory control policies should generate replenishment orders guaranteeing full truckloads, resulting in low costs. In this paper two replenishment policies, satisfying this requirement, are considered. We start our discussion with the QS policy (Renberg and Planche (1967)).

The QS policy is an aggregate continuous review order policy where the aggregate inventory position (the sum of all individual inventory positions defined as stock on hand plus stock on order minus backorder) triggers an order whenever it drops below a reorder level s . Then the individual inventory positions of each item are reviewed and an order-up-to policy is applied, which means an order for item i is placed equal to the difference between the order-up-to level S_i and the actual inventory position of item i . If Q denotes the capacity of a vehicle and N the number of different items, full truckloads are obtained under unit demand sizes and the following condition:

$$\sum_{i=1}^N S_i - s = Q \quad (1)$$

Since we allow non-unit demand sizes the actual aggregate inventory position can be less than the reorder level at the moment when an order is triggered. We rely on approximations to compute the policy parameters and therefore, we also neglect the undershoot in the formulas. However, in the simulations we assume that for the item, causing the undershoot, the order is split and the remaining part has to wait until the next truck is leaving. We further assume a single location where all items are kept on stock and one sup-

plier location, resulting in a common replenishment leadtime L for all items. Additional, no routing decisions have to be made.

The demand process of each item is modeled as a compound renewal process which means that demand sizes D_i as well as interarrival times of demand A_i are stochastic. We assume that interarrival times as well as demand sizes are distributed according to a mixture of two Erlang distributions $\mathbf{E}_{k_1, k_2}((\mu_1, \mu_2), (p_1, p_2))$ and that the first two moments are known. The density of an $\mathbf{E}_{k_1, k_2}((\mu_1, \mu_2), (p_1, p_2))$ distribution is given by

$$f_X(x) := \sum_{i=1}^2 p_i \mu_i^{k_i} \frac{x^{k_i-1}}{(k_i-1)!} e^{-\mu_i x} \quad x > 0 \quad (2)$$

Unsatisfied demand is backordered and for each item a target service level β_i^{target} is given. Here we consider the fill rate, defined as the percentage of demand which can be directly satisfied from stock. In the following we will show how to compute the $N + 1$ policy parameters $(s, S_1, S_2, \dots, S_{N-1}, S_N)$ such that the target fill rates can be met.

Below we summarize the definition of the parameters and variables and the used notation:

N :	Number of different items
A_i :	Time between two subsequent arrivals of demand of item i
D_i :	Demand size of item i
A^* :	Time between two subsequent arbitrary arrivals of demand
D^* :	Demand size of an arbitrary demand
$D_i(T)$:	Demand of item i during an interval with length T
β_i^{target} :	Target service level of item i
Q :	Capacity of the vehicle
S_i :	Order-up-to level for item i
S_0 :	Total order-up-to level ($S_0 := \sum_{i=1}^N S_i$)
s :	Reorder level
L :	Replenishment leadtime
σ :	Length of a replenishment cycle
$VAR[X]$:	Variance of a random variable X
c_X^2 :	Squared coefficient of variation of a random variable X ($c_X^2 = \frac{VAR[X]}{E[X]}$).

2.2 Replenishment cycle

The length of a replenishment cycle σ is random and determined by the aggregate demand process (A^*, D^*) which is the superposition of the individual demand processes (A_i, D_i) . Assuming that the aggregate demand process can also be modeled as a compound renewal process, approximations for the first two moments of the inter-arrival time A^* can be computed using the stationary interval method developed by Whitt (1982) to superpose renewal processes. Instead of superposing hyper-exponential and shifted exponential distributions, we superpose mixtures of Erlang distributions (for a detailed algorithm see Appendix I). The first two moments of the aggregated demand sizes can be computed as weighted sum of the individual demand sizes as follows:

$$E[D^*] = E[A^*] \sum_{i=1}^N \frac{E[D_i]}{E[A_i]} \quad (3)$$

$$E[(D^*)^2] = E[A^*] \sum_{i=1}^N \frac{E[(D_i)^2]}{E[A_i]} \quad (4)$$

The length of a replenishment cycle σ is given by

$$\sigma = \sum_{j=1}^{N(S_0-s)} A_j^* \quad (5)$$

where A_j^* denotes the interarrival time between arrival $j - 1$ and j and $N(S_0 - s)$ denotes the number of arrivals of arbitrary demands between two replenishments. It is well known that the first two moments of the length of the replenishment cycle as given in (5) can be computed as follows:

$$E[\sigma] = E[N(S_0 - s)]E[A^*] \quad (6)$$

$$E[\sigma^2] = E[N(S_0 - s)]VAR[A^*] + E[N^2(S_0 - s)]E^2[A^*] \quad (7)$$

Note that the moments for the aggregate interarrival-time are approximated and therefore, the numerical values of $E[\sigma]$ and $E[\sigma^2]$ will also be approximations although (6) and (7) are exact formulas. Moreover, we also approximate the moments for the number of arrivals of arbitrary demands between two replenishments, because especially an exact

expression for the second moment of $N(S_0 - s)$ is in general intractable. First, we assume that the undershoot is small compared to the truck capacity and can therefore be neglected. This leads to the following relation:

$$S_0 - s \approx \sum_{j=1}^{N(S_0-s)} D_j^* \quad (8)$$

where D_j^* denotes the arbitrary demand size of the j -th arrival. We obtain for the first moment:

$$E[N(S_0 - s)] \approx \frac{S_0 - s}{E[D^*]} \quad (9)$$

For the second moment we additionally rely on approximations based on renewal theory (Cox (1962)).

$$E[N^2(S_0 - s)] \approx \frac{(S_0 - s)^2}{E^2[D^*]} + (S_0 - s) \frac{c_{D^*}^2}{E[D^*]} + \frac{E^2[(D^*)^2]}{2E^4[D^*]} - \frac{E[(D^*)^3]}{3E^3[D^*]} \quad (10)$$

In order to get reasonable results the vehicle capacity Q must be large enough compared to the aggregate demand size (cf. Tijms (1994)), which means $Q > Cond(D^*)$ with

$$Cond(D^*) := \begin{cases} \frac{3}{2}c_{D^*}^2 E[D^*] & : c_{D^*}^2 > 1 \\ E[D^*] & : 0.2 < c_{D^*}^2 \leq 1 \\ \frac{1}{2c_{D^*}} E[D^*] & : 0 < c_{D^*}^2 \leq 0.2 \end{cases} \quad (11)$$

2.3 The service level

The fill rate for a specific item i , ($i = 1, 2, \dots, N$) is given by the following formula (cf. Tijms Tijms (1994))

$$\beta_i^{target} = 1 - \frac{E[(D_i(L + \sigma) - S_i)^+] - E[(D_i(L) - S_i)^+]}{E[D_i(\sigma)]} \quad (12)$$

Under the assumption that a random variable X is distributed according to a mixture of two Erlang distributions $\mathbf{E}_{k_1, k_2}((\mu_1, \mu_2), (p_1, p_2))$ the following formula holds.

$$\begin{aligned} E[(X - z)^+] &= \int_z^\infty (x - z) dF_X(x) \\ &= \sum_{j=0}^1 (-z)^j \left(\sum_{i=1}^2 p_i \frac{\mu_i^{k_i}}{(k_i - 1)!} \sum_{l=0}^{k_i-j} \frac{z^l (k_i - j)!}{l! \mu_i^{k_j - l - j + 1}} e^{-\mu_i z} \right) \end{aligned} \quad (13)$$

Therefore, the fill rate (12) can easily be computed under the assumption that the demand during an interval with length T is mixed Erlang distributed. We compute the first two moments of the demands during the intervals with length σ , L , and $L + \sigma$, fit a mixed-Erlang distribution on these moments and use (12) as an approximation for the fill rate. How to fit a mixed Erlang distribution on the the first two moments of a random variable is described in Appendix III.

In order to obtain the first two moments of the demand during an interval with length T we use

$$E[D_i(T)] = E[N_i(T)]E[D_i] \quad (14)$$

$$E[D_i^2(T)] = E[N_i(T)]VAR[D_i] + E[N_i^2(T)]E^2[D_i] \quad (15)$$

and we again rely on approximations of renewal theory as follows:

$$E[N_i(T)] \approx \frac{E[T]}{E[A_i]} \quad (16)$$

$$E[N_i^2(T)] \approx \frac{E[T^2]}{E^2[A_i]} + E[T] \left(\frac{E[A_i^2]}{E^3[A_i]} - \frac{1}{E[A_i]} \right) + \frac{E^2[A_i^2]}{2E^4[A_i]} - \frac{E[A_i^3]}{3E^3[A_i]} \quad (17)$$

The formulas (16) and (17) lead to reasonable results if $P(T < Cond(A_i)) \leq \epsilon$ for a small value of ϵ . Otherwise, when the length of the interval is small compared to the interarrival times, the number of arrivals during this interval is small and we can compute the distribution function of $N(T)$ numerically. Using the distribution function the moments are computed according to their definitions (see Appendix II).

2.4 Computation of the policy parameters

Based on the formulas obtained above the policy parameters $(s, S_1, S_2, \dots, S_N)$ for the QS-policy can be computed such that a required service level is obtained. As input for the following algorithms numerical values for the vehicle capacity Q , the leadtime L , the target service levels β_i^{target} , and the first two moments of the demand sizes D_i and the interarrival times A_i are needed. A sketch of the algorithm is given in the following.

1. Compute the moments of the aggregate demand process (A^*, D^*) using equation (3) and (4) and the algorithm presented in Appendix I.
2. Set $Q = S_0 - s$
3. Compute $E[N(S_0 - s)]$ and $E[N^2(S_0 - s)]$ according to (9) and (10)
4. Compute the moments of the replenishment cycle $E[\sigma]$ and $E[\sigma^2]$ using (6) and (7)
5. Set $T = \sigma$ in (14) and (16) and compute $E[D_i(\sigma)]$ for all $i = 1, 2, \dots, N$
6. For $T = L$ and $T = L + \sigma$ compute the first two moments of $D_i(T)$ for all $i = 1, 2, \dots, N$ using (14) - (17)
7. For each $i = 1, 2, \dots, N$ fit a mixed Erlang distribution and compute the parameters p and k as described in Appendix III
8. Use a bisection method and (12) combined with (13) in order to compute the policy parameters S_i for each $i = 1, 2, \dots, N$.
9. Set $s = \sum_{i=1}^N S_i - Q$

2.5 Quality of the approximations

Since most of the formulas obtained above are approximations we have to test the quality of them. In order to get detailed insights an extensive simulation study is performed. For given target service levels the policy parameters are computed using our provided approximations and afterwards the system is simulated and the difference between the target service level and the actual service level is measured. Since we allow non-unit demand sizes the actual aggregate inventory position can be less than the reorder level at the moment when an order is triggered. In the simulation we assume that for the item, causing the undershoot, the order is split and the remaining part has to wait until the next truck is leaving.

For each parameter set 10 simulation runs with different seeds are performed yielding accurate point estimates for the measured performance characteristics. For each run the

actual service level and the absolute deviation from the target service level is computed for all items and afterwards the average over all runs and items is computed as follows:

$$\Delta_{av} := \frac{1}{10N} \sum_{j=1}^{10} \sum_{i=1}^N | \beta_{i,j}^{sim} - \beta_i^{target} | \quad (18)$$

where $\beta_{i,j}^{sim}$ denotes the actual service level of item i in simulation run j .

Additionally, we compute the maximum positive and negative deviation between target and simulated service level over all items and runs:

$$\Delta_{max} := \max_{i,j} \{ (\beta_{i,j}^{sim} - \beta_i^{target})^+ \}, \quad \Delta_{min} := \max_{i,j} \{ (\beta_i^{target} - \beta_{i,j}^{sim})^+ \} \quad (19)$$

where $(x)^+$ is defined as $\max\{0, x\}$.

To test the approximations we have chosen the following parameter values as a base case:

$$N = 32, \quad Q = 5000, \quad L = 2, \quad \beta_i^{target} = 0.95 \quad \forall i = 1, 2, \dots, N \quad (20)$$

Further, the average demand size is uniformly drawn from the interval (10,60) and the average interarrival time of orders is uniformly drawn from the interval (0.1,1.1). Additionally, we consider 9 different examples related to the coefficient of variation of the demand sizes and the interarrival times $c_A^2, c_D^2 \in \{0.4, 1.0, 1.6\}$.

We tested the impact of the parameters on the performance of the approximation, varying one parameter while keeping the others fixed as in the base case.

We could not observe any impact of the replenishment leadtime on the performance of the approximations ($L \in \{2, 4, 6, 8, 10\}$). Therefore, the leadtime is fixed to $L = 2$ in the study below. We further investigated five different target values ($\beta^{target} \in \{0.75, 0.85, 0.9, 0.95, 0.99\}$). For larger target service levels smaller deviations can be observed than for smaller target levels (see Figure 1 where the average absolute deviation is illustrated for different values of (c_A^2, c_D^2) and identical as well as non identical items).

We can further observe that for small squared coefficient of variation of the interarrival times the fill rate is underestimated by (12) while for a large variability the fillrate is overestimated. This is illustrated in Table 1 where the maximum positive and the maximum negative deviation of the service level are given for a capacity $Q = 500$ and $\beta^{target} = 0.95$.

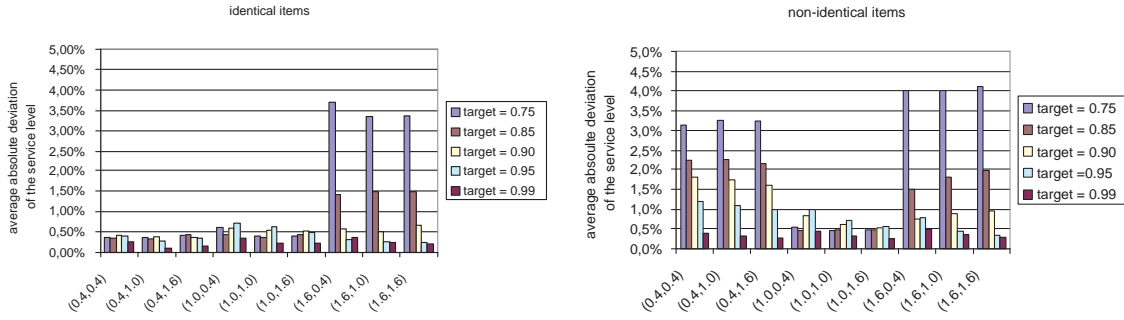


Figure 1: Impact of the target service level on the approximations

(c_A^2, c_D^2)	Δ_{max}	Δ_{min}	(c_A^2, c_D^2)	Δ_{max}	Δ_{min}	(c_A^2, c_D^2)	Δ_{max}	Δ_{min}
(0.4,0.4)	0.04	0.00	(1.0,0.4)	0.01	0.00	(1.6,0.4)	0.00	0.02
(0.4,1.0)	0.04	0.00	(1.0,1.0)	0.01	0.00	(1.6,1.0)	0.00	0.01
(0.4,1.6)	0.04	0.00	(1.0,1.6)	0.01	0.00	(1.6,1.6)	0.00	0.01

Table 1: Maximum positive and maximum negative deviation of the service level

However, the number of items and the capacity of the vehicle have the largest impact on the quality of the approximations. The following section is devoted to a study of this effect.

2.5.1 The impact of the number of items

For our approximations we determine the aggregate demand process by superposing the individual demand processes. It is well known that the number of superposed processes has an impact on the accuracy of the moments, especially on the second moment of the superposed process. Therefore, we report below on the impact of the number of items on the approximations for the nine different examples related to (c_A^2, c_D^2) . The left side of Figure 2 illustrates the results for items with identical parameter values and on the right side the results for non-identical parameters are illustrated. For the Figure the target service level is chosen $\beta^{target} = 0.95$ for all items, and results are depicted for different combinations of (c_A^2, c_D^2) .

It can be seen that the deviations from the target service level are in general quite small and that the quality of the approximation is improving with increasing number of items.

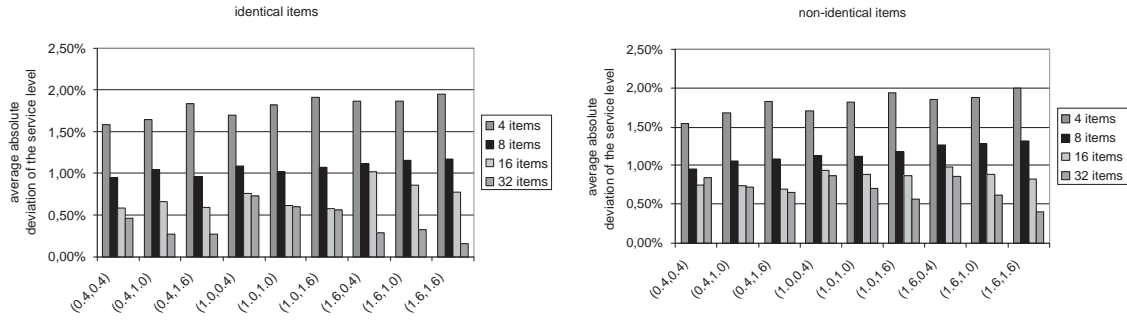


Figure 2: Impact of the number of items

Further, there is not a significant difference between the identical and non-identical case. Moreover, the impact of the coefficient of variation of the demand size and the interarrival time is limited. The target service level is always underestimated by (12), since $\Delta_{min} = 0$.

2.5.2 The impact of the vehicle capacity

It is already mentioned that for getting reasonable results with relation (10) it is necessary that the vehicle capacity is large enough compared to the average aggregate demand size. The impact of the vehicle capacity is illustrated in Figure 3 for different values of (c_A^2, c_D^2) .

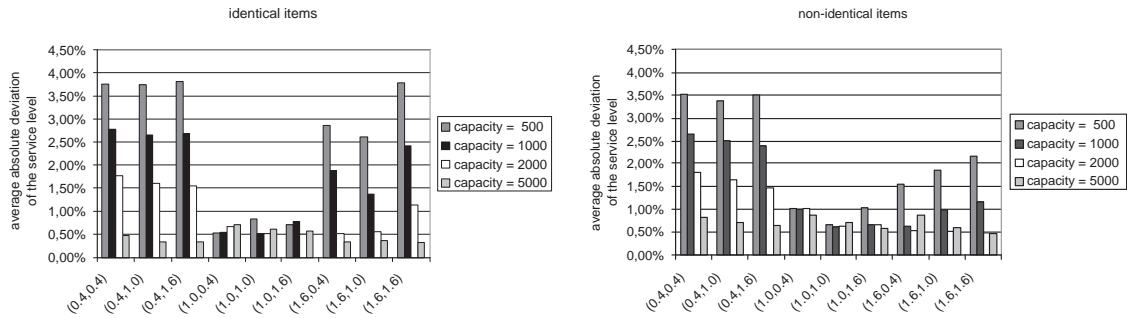


Figure 3: Impact of the vehicle capacity

It can be seen that the capacity of the vehicle has not much impact on the quality of the approximations when the demand arrivals follow a Poisson process. For the other examples the performance is improving with increasing vehicle capacity. Based on our simulation study we can conclude that the approximations work well as long as the number

of items and the vehicle capacity are large enough. In the following we will use the approximations and we will compare the QS policy with a policy where replenishment orders are triggered by individual items.

3 Individual order triggering

Since under the control of a QS policy replenishment orders are only triggered when the aggregate inventory drops below the reorder level, the replenishment opportunity for an item is dependent from the demand process of all other items. Therefore, there is a chance that an item is running out of stock and no order is placed or, in order to avoid such a situation, high safety stocks are needed. Therefore, we compare in the following the performance of the QS policy with a policy where replenishment orders are triggered by the inventory position of an individual item. Whenever the inventory position of an individual item drops below the item specific reorder level a replenishment order is triggered. Then the total order quantity, which is assumed to be equal to a full truckload, has to be allocated among all items. A reasonable criteria for the allocation procedure is the expected elapsed time until the next replenishment order is triggered (Miltenburg (1985)), in the sequel also called the runout-time. In order to determine the allocation which maximizes the expected time until the next replenishment order is triggered, the following optimization problem has to be solved

$$\begin{aligned} & \max_{q_1, \dots, q_N} \alpha \\ \text{s.t.} & \quad (IP_i - s_i + q_i) \cdot \frac{E[A_i]}{E[D_i]} \geq \alpha \quad \forall i = 1, 2, \dots, N \\ & \quad \sum_{i=1}^N q_i = Q \end{aligned}$$

where IP_i denotes the actual inventory position of item i . In order to solve this problem we rely on the following method. In the first step the runout time $E[\tau_i]$ is computed for all items $i = 1, 2, \dots, N$ by:

$$E[\tau_i] = E[A_i] \cdot \frac{(IP_i - s_i)}{E[D_i]} \quad (21)$$

In the second step the items are ordered according to their runout times. Without loss of generality we assume that item one has the shortest runout time and that item N

has the largest.

$$E[\tau_1] \leq E[\tau_2] \leq \dots \leq E[\tau_N] \quad (22)$$

The third step is composed of several loops.

- 1) Set $i_{max} := 1$
- 2) Increase all ordersizes q_i for $i \leq i_{max}$ until the runout times are equal for all $i \leq i_{max} + 1$ or a full truckload is achieved.
- 3) If $i_{max} < N$ then set $i_{max} := i_{max} + 1$ and repeat step 2 and 3.

This process stops when the total order quantity is allocated or the runout times of all items are equal. In the latter case allocation is continued and order sizes are enlarged simultaneously while keeping the runout times equal until the total order quantity is allocated. As a result of the allocation algorithm all replenishment orders can be shipped as full truckloads and the expected time until the next order is placed is maximized.

3.1 Numerical comparison of the policies

In order to compare the QS policy (aggregate order triggering) with the allocation policy (individual order triggering) a simulation study is done. Since no easy formulas for the allocation policy are available to compute the reorder levels (s_1, \dots, s_N) such that a given target service level can be met, we first simulate the allocation policy and then used the measured service level for each item as the target service levels for the QS policy. Due to the simulation and the used approximations the actual service level of the QS policy can differ from the target service level. Therefore, we measure differences of the average inventory as a function of the deviation from the service level.

3.1.1 Identical items

We have measured the difference between the service levels of both policies for item i :

$$\Delta_i := \frac{1}{10} \sum_{j=1}^{10} \beta_{ij}^{QS} - \frac{1}{10} \sum_{j=1}^{10} \beta_{ij}^{allo} \quad (23)$$

with $\beta_{i,j}^{QS}$ denoting the service level for item i in simulation run j for the corresponding QS policy and with $\beta_{i,j}^{allo}$ denoting the service level for item i in simulation run j for the allocation policy. Additionally we measure the relative difference of the average inventory of item i defined as

$$\delta_i := \frac{I_i^{QS} - I_i^{allo}}{I_i^{QS}} \cdot 100\% \quad (24)$$

where

$$I_i^{QS} := \frac{1}{10} \sum_{j=1}^{10} E[I_{i,j}^{QS}] \quad (25)$$

and $E[I_{i,j}^{QS}]$ is defined as the average inventory of item i in simulation run j for the QS policy. Although the average number of orders is the same, the average number of items included in an order may be different for both policies. We additionally measure M_{QS} (M_{allo}), the average number of items included in an order under the QS (allocation) policy.

Below we report on the results of 117 different examples with 16 identical items with respect to the stochastic parameters. For the coefficients of variations we have chosen $c_A^2, c_D^2 \in \{0.4, 1.0, 1.6\}$. For Figure 4 we have fixed $E[A] = 0.1$ and $E[D] = 50$ ($E[D] = 20$) and we varied the service level ($\beta_i^{target} \in \{0.82, 0.89, 0.93, 0.98\}$). In another series of examples we fixed $\beta_i^{target} = 0.93$ and we varied the average interarrival time ($E[A] \in \{0.1, 0.5, 1.0, 2.0, 5.0\}$).

For each item we depict in Figure 4 the relative deviation of the inventory as a function of the deviation of the service level for different target service levels.

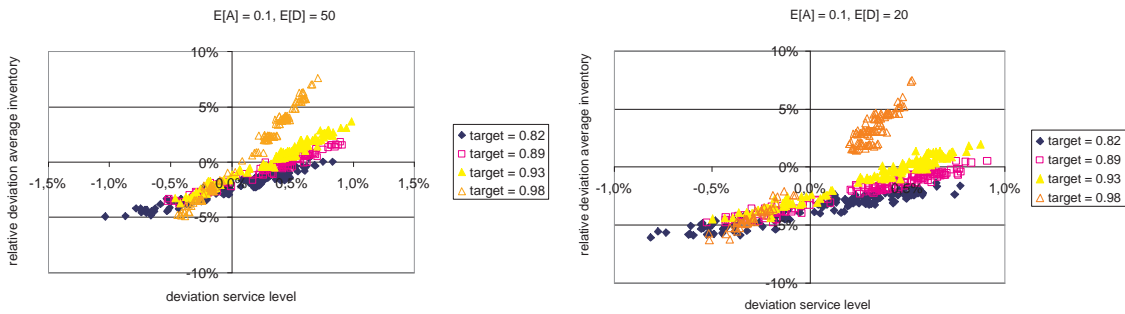


Figure 4: Comparison of the policies (identical items)

For different average interarrival times the results are illustrated in Figure 5.

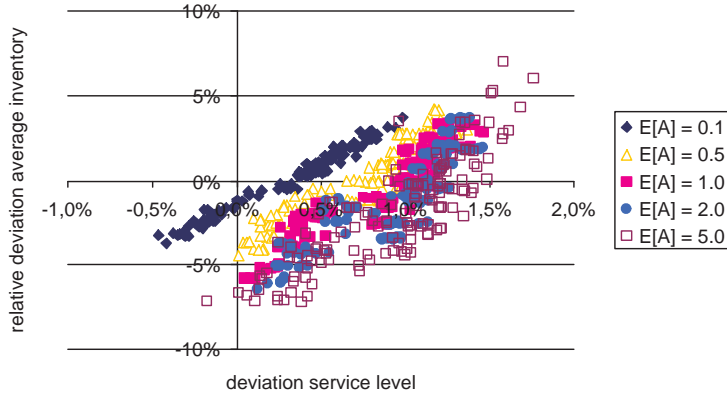


Figure 5: Comparison of the policies (identical items)

It can be seen that deviations are different for different target service levels as well as for different interarrival times. The different coefficient of variations do not seem to have much impact on the relation between servicelevel deviation and inventory deviation. Based on our numerical experiments we build up regression models in order to estimate the relative difference in inventory for the case of equal service levels. The results are summarized in Table 2.

$E[A]$	$E[D]$	β^{target}	δ	$E[A]$	$E[D]$	β^{target}	δ	$E[A]$	$E[D]$	β^{target}	δ
0.1	50	0.82	-2.30	0.1	20	0.82	-3.74	0.1	50	0.93	-1.47
0.1	50	0.89	-1.85	0.1	20	0.89	-3.14	0.5	50	0.93	-4.05
0.1	50	0.93	-1.47	0.1	20	0.93	-2.47	1.0	50	0.93	-5.71
0.1	50	0.98	-0.65	0.1	20	0.98	-0.64	2.0	50	0.93	-6.69

Table 2: Comparison of the policies (identical items)

It can be seen that in all cases more inventory is needed for the allocation policy to get the same service level as with the QS policy although the differences are sometimes small. Especially for slow moving items with large average interarrival time and small ordersizes the QS policy outperforms the allocation policy. However, we have to mention that the average number of items included in an order is less in case of the allocation policy (see Figure 6). But as long as the variability of the demand size is not too large, these differences are negligible.

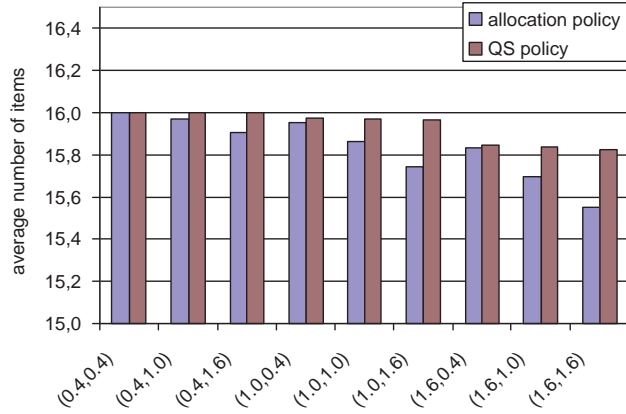


Figure 6: Average number of different items per order

3.1.2 Non-identical items

In case of non-identical items the parameters are uniformly drawn from different intervals as follows:

Examples 1: $E[A] \sim (0.1, 1.0)$ $E[D] \sim (20, 40)$

Examples 2: $E[A] \sim (0.1, 1.0)$ $E[D] \sim (50, 100)$

Examples 3: $E[A] \sim (0.1, 2.0)$ $E[D] \sim (20, 40)$

Examples 4: $E[A] \sim (0.1, 2.0)$ $E[D] \sim (50, 100)$

Examples 5: $E[A] \sim (1.0, 3.0)$ $E[D] \sim (20, 40)$

For each example we consider 16 items and nine different combinations for the coefficient of variations ($c_A^2, c_D^2 \in \{0.4, 1.0, 1.6\}$). We further set the parameter values such that reorder levels for the allocation policy result in fill rates between 90% - 98% where fast movers in general have larger service levels than slow movers.

Since in all examples almost all items have a larger service level in case of the QS policy compared to the allocation policy we only illustrate the relative deviation of the inventory levels. We observe a large impact of the average interarrival time of orders on changes in inventory levels. This is also illustrated in Figure 7 where we depict the

results of all 16 items for different demand characteristics. On the left-hand side demand variability is low ($c_A^2, c_D^2) = (0.4, 0.4)$) while on the right-hand side the demand variability is large ($c_A^2, c_D^2) = (1.6, 1.6)$).

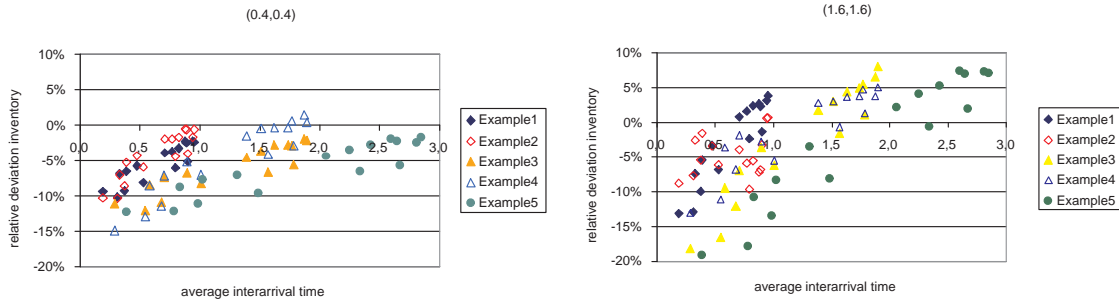


Figure 7: Comparison of the policies (non-identical items)

It can be seen that especially fast moving items need much less inventory to satisfy a required service level under a QS policy compared with the allocation policy. If the variability of the demand process is not too large all items benefit from the QS policy, while for large demand variability some slow moving items will need more inventory. Note that part of the increase of the inventory level may also be induced by an increased service level and that the fast moving items also have larger average inventory levels. For example, the item with a relative deviation of the inventory level of 19% leads to 218,58 units less on stock while the item with an increase of 7% leads to 12.52 units more on stock (Example 5, right side Figure 7). In general the reduction of inventory levels for fast moving items is much larger than for slow moving items, resulting in a reduced total inventory level in case of the QS policy. Moreover, the differences between total inventory levels for both policies is dependent on the ration between fast and slow movers. We illustrate this effect by means of an example where the average interarrival times of slow movers are set equal to 2.0 and the average interarrival time of fast movers is set equal to 0.2. For a fixed number of items (16) we vary the ratio between fast and slow movers and measure the relative difference of the total average inventory levels for the group of the fast and the slow movers separately. The results are depicted in Figure 8 for different coefficient of variations.

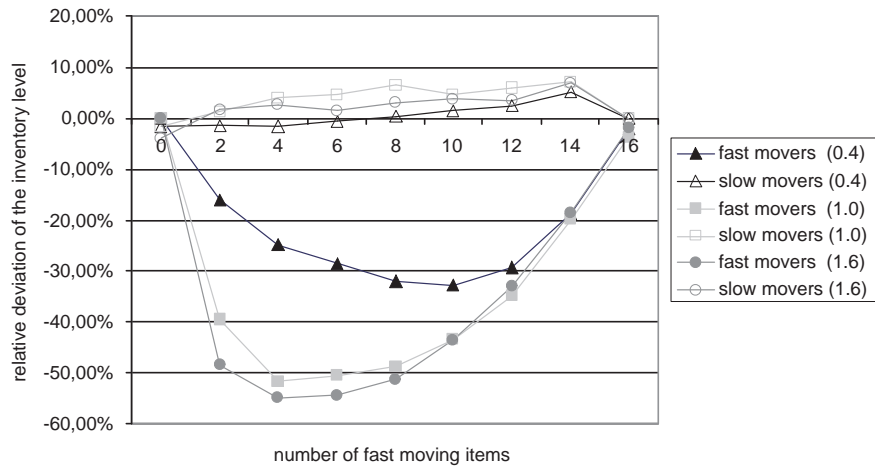


Figure 8: The impact of ratio between fast and slow movers

The impact of the ratio between fast and slow movers and the different behavior of them is obvious. The slightly increase of the inventory levels of the slow movers cannot overrule the strong decrease of the inventory levels of the fast movers, resulting in lower total inventory under the QS policy. Moreover, effects are larger under more variability of the item demand.

For an explanation of the effects we have to consider the replenishment process. While the average length of an replenishment cycle is the same for both policies, the variance of the replenishment cycle is different. In case of individual order triggering we can observe much more variability in the replenishment moments which requires more safety stock, compared to aggregate order triggering, to achieve the same service level.

However, we have to mention that we have only compared inventory levels and not inventory holding costs. If slow and fast movers differ substantially in holding costs (keeping a slow moving item on stock must be 50 times more expensive than keeping a fast moving item on stock), then it may happen that the total inventory costs may not decrease. Similar, total cost may not decrease, when line item costs are not small compared to fixed order costs.

4 Summary and conclusions

In this paper we have provided approximated formulas to compute the parameters of an QS policy under compound renewal demand such that given target service levels can be met. We have tested the performance of the approximations and based on a numerical simulation study we can conclude that the approximations are accurate as long as the number of items and the vehicle capacity is not too small.

Additionally, we have illustrated that under a service level approach the QS policy has smaller average inventory levels than the allocation policy, since the allocation policy creates much more variability in the replenishment process and therefore, higher safety stocks are needed. However, the difference is influenced by the demand characteristics of the items. While for identical items the difference between both policies is negligible, large differences can be observed under different item characteristics. The impact on total cost is dependent on the costs parameters, but as long as fixed order costs are dominant and holding costs of items do not differ substantially, the QS policy outperforms the allocation policy.

References

- Atkins, D.R., Iyogun, P.O., 1988. Periodic versus can-order policies for coordinated multi-item inventory systems, *Management Science* 34 (6) 791–796.
- Aksoy, Y., Selcuk Erenguc, S., 1987. Multi-Item Inventory Models with Co-ordinated Replenishments: A Survey, *International Journal of Operations Management* 8 (1) 63–73.
- Balintfy, J.L., 1964. On a Basic Class of Multi-Item Inventory Problems, *Management Science* 10 (2) 287–297.
- Carlson, M.L., Miltenburg, G.J., 1988. Using the Service Point Model to Control Large Group of Items, *Omega* 16 (5) 481–489.
- Cox, D.R., 1962. *Renewal theory*, Methuen, London.
- Federgruen, A., Groenevelt, H., Tijms, H.C., 1984. Coordinated replenishments in a multi-item inventory system with compound Poisson demands, *Management Science* 30 (3) 344–357.
- Fung, R.Y.K., Ma, X., Lau H.C.W., 2001. (T,S) policy for coordinated inventory replenishment systems under compound Poisson demands, *Production Planning and Control* 12 (6) 575–583.
- Goyal, S.K., Satir, A.T., 1989. Joint replenishment inventory control: Deterministic and Stochastic models, *European Journal of Operational Research* 38 2–13.
- Ignall, E., 1969. Optimal continuous review policies for two product inventory systems with joint setup costs, *Management Science* 15 (5) 278–283.
- Johansen, S.G., Melchior P., 2003. Can-order policy for the periodic-review joint replenishment problem, *Journal of the Operational Research Society* 54 283–290.
- Melchior, P., 2002. Calculating can-order policies for the joint replenishment problem by the compensation approach, *European Journal of Operational Research* 141 587 – 595.

- Miltenburg, G.J., 1985. Allocating a Replenishment Order Among a Family of Items, *IIE Transactions* 17 (3) 261 – 267.
- Pantumsinchai, P., 1992. A comparison of three joint ordering inventory policies, *International Journal of Decision Science* 23 (1) 111–127.
- Renberg, B., Planche, R., 1967. Un Modele Pour La Gestion Simultanee Des n Articles D'un Stock, *Revue Francaise d'Informatique et de Recherche Operationelle* 6 47–59.
- Schultz, H., Johansen, S.G., 1999. Can-order policies for coordinated inventory replenishment with Erlang distributed times between ordering, *European Journal of Operational Research* 13 30–41.
- Silver, E.A., 1974. A control system for coordinated inventory replenishment, *International Journal of Production Research* 12 (6) 647–671.
- Silver, E.A., 1981. Establishing reorder points in the (S,c,s) coordinated control system under compound Poisson demand, *International Journal of Production Research* 19 (6) 743–750.
- Swenseth, S.R., Godfrey, M.R., 2002. Incorporating transportation costs into inventory replenishment decisions, *International Journal of Production Economics* 77 113–130.
- Thompstone, R., Silver, E.A., 1975. A coordinated inventory control system for compound Poisson demand and zero leadtime, *International Journal of Production Research* 13 (6) 581–602.
- Tijms, H.C., 1994. *Stochastic models: an algorithmic approach*, John Wiley & Sons.
- Van Eijs, M.J.G., 1994. On the determination of the Control Parameters of the Optimal Can-Order Policy, *Zeitschrift für Operations Research* 39 289–304.
- Viswanathan, S., 1997. Note, Periodic Review (s,S) Policies for Joint Replenishment Inventory Systems, *Management Science* 43 (10) 1447–1454.
- Whitt, W., 1982. Approximating a point process by a renewal process I: Two basic methods, *Operations Research* 30 125–147.

Appendix I: Algorithm to compute the first two moments of a superposed compound renewal process

The inter-arrival time of the single processes is denoted with X_i and of the superposed process it is denoted with X_0 . The following iterative algorithm can be used to compute the first two moments of X_0 .

1. Order the N inter-arrival processes from the largest to the smallest first moment.
2. Compute

$$E[X_0^{(1)}] := \frac{1}{\sum_{i=1}^2 \frac{1}{E[X_i]}} \quad (26)$$

$$E[(X_0^{(1)})^2] := 2E[X_0^{(1)}] \int_0^\infty \left(\prod_{i=1}^2 \frac{1}{E[X_i]} \right) \left(\prod_{i=1}^2 \int_x^\infty (1 - F_{X_i}(y)) dy \right) dx \quad (27)$$

3. Fit a mixed-Erlang distribution to the first two moments of $X_0^{(1)}$.
4. Initially set $n:=2$ and $i:=3$.
5. Compute

$$E[X_0^{(n)}] := \frac{1}{\frac{1}{E[X_0^{(n-1)]}} + \frac{1}{E[X_i]}} \quad (28)$$

$$E[(X_0^{(n)})^2] := 2E[X_0^{(n)}] \int_0^\infty \frac{1}{E[X_0^{(n-1)]}E[X_i]} \left(\int_x^\infty (1 - F_{X_0^{(n-1)}}(y)) dy \int_x^\infty (1 - F_{X_i}(y)) dy \right) dx \quad (29)$$

Fit a mixed-Erlang distribution to the first two moments of $X_0^{(n)}$

6. If $n < N$ then $n := n + 1$, $i := i + 1$ and go to step 5 else $E[X_0] := E[X_0^{(N)}]$ and $E[X_0^2] := E[(X_0^{(N)})^2]$

Appendix II: Number of arrivals during an interval with length T

The following relation holds for the probability function of $N(T)$:

$$P(N(T) = k) = P(N(T) \geq k) - P(N(T) \geq k + 1) \quad (30)$$

We obtain

$$P(N(T) = k) := \begin{cases} P(A_1 \geq T) & : k = 0 \\ P(\sum_{j=1}^k A_j \leq T) - P(\sum_{j=1}^{k+1} A_j \leq T) & : k = 1, 2, \dots, k_{max} \end{cases} \quad (31)$$

where A_j denotes the interarrival time between arrival $j - 1$ and j and k_{max} is chosen such that

$$\sum_{j=1}^{k_{max}} P(N(T) = j) \geq 0.9999 \quad (32)$$

For the computation of $P(\sum_{j=1}^k A_j \leq T)$ we rely on a two moment fit of a mixed Erlang distribution.

Appendix III: How to fit a mixed-Erlang distribution

The density of an $\mathbf{E}_{k_1, k_2}((\mu_1, \mu_2), (p_1, p_2))$ distribution is given by

$$f_X(x) := \sum_{i=1}^2 p_i \mu_i^{k_i} \frac{x^{k_i-1}}{(k_i-1)!} e^{-\mu_i x} \quad x > 0 \quad (33)$$

Assume the first two moments of X to be given. The parameters p_i , k_i and μ_i ($i = 1, 2$) can be found from the first two moments of X as follows:

- If $c_X^2 < 1$ then a mixture of two Erlang distributions with the same scale parameter is used. Hence,

$$k_1 = \lfloor \frac{1}{c_X^2} \rfloor$$

$$k_2 = k_1 + 1$$

$$p_1 = \frac{1}{1+c_X^2} \left(k_2 c_X^2 - \sqrt{k_2(1+c_X^2) - k_2^2 c_X^2} \right)$$

$$p_2 = 1 - p_1$$

$$\mu_1 = \frac{k_2 - p_1}{E[X]}$$

$$\mu_2 = \mu_1$$

- If $c_X^2 \geq 1$ then a mixture of two exponential distributions used.

$$k_1 = 1$$

$$k_2 = 1$$

$$\mu_1 = \frac{2}{E[X]} \left(1 + \sqrt{\frac{c_X^2 - \frac{1}{2}}{c_X^2 + 1}} \right)$$

$$\mu_2 = \frac{4}{E[X]} - \mu_1$$

$$p_1 = \frac{\mu_1(\mu_2 E[X] - 1)}{\mu_2 - \mu_1}$$

$$p_2 = 1 - p_1$$