Approximate evaluation of order fill rates for an inventory system of service tools

I.M.H.Vliegen* G.J. van Houtum
i.m.h.vliegen@tm.tue.nl g.j.v.houtum@tm.tue.nl

Department of Technology Management, Technische Universiteit Eindhoven, P.O. Box 513, 5600 MB Eindhoven, the Netherlands

February 28, 2007

Abstract

This paper deals with the analysis of a single-location, multi-item inventory model for service tools, in which coupled demands and coupled returns occur. We distinguish multiple Poisson demand streams. Per stream there is a given set of tools that is requested per demand. We are interested in the order fill rates, i.e., the percentage of demands for which all requested tools are delivered from stock. Requested tools that are not on stock are delivered via an emergency channel. For the warehouse under consideration, they may be considered as lost sales. Delivered tools are returned to the warehouse after a deterministic return time, that is equal for all tools. We develop three approximate models for the order fill rates, which are all based on Markovian models. One approximate model has appeared to give an underestimation in all computational tests, while the second approximate model has led to an overestimation in all instances tested. The last approximate model combines the other two. This approximate model is very accurate and can be computed efficiently. Hence, it is appropriate for usage in multi-item optimization algorithms.

Keywords: Single stock point, multiple service tools, coupled demands, coupled returns, order fill rate.

1. Introduction

Original Equipment Manufacturers (OEMs) produce expensive machines, that are critical in the production process of their customers. Therefore, customers often have service contracts with the OEM in which the availability of the machine is agreed upon. To make sure this performance is met, the OEM performs preventive maintenance. Furthermore, in case of a defect, corrective maintenance is performed, for which the company needs spare parts, service engineers, and service tools. These resources are positioned in a global network consisting of central and local customer

* *Corresponding author*

service points. The company's objective is to meet the agreed system performance against minimal costs. Total costs consist of procurement costs, inventory holding costs, transportation costs for regular, lateral and emergency shipments of spare parts and service tools, import taxes, and the costs for employing service engineers.

The OEM has to decide how much stock is needed at which location, both for spare parts and service tools, and how many service engineers should be hired to meet all service targets. In practice, the system performance demanded by customers is decoupled into separate targets for engineers, spare parts and service tools, and the minimization problem is solved for each resource separately. In this paper we focus on the subproblem involving service tools. So far, the stock planning of service tools has received only little attention in literature. However, since prices of service tools can be very high, and service tools may lead to large investments for OEM's, optimization of stock levels of service tools is an important issue.

The situation described occurs for many OEMs, among which a company in the semiconductor supplier industry, with whom we collaborate. This company already uses sophisticated methods to optimize stock levels for spare parts. For service tools however the company is still in need of a model to determine the optimal stock levels. The service tools at this company are stored at local warehouses. When tools are needed for a maintenance action, they are taken from the nearest warehouse by a service engineer, and after usage they are returned to this warehouse. Service tools can also be lent to other warehouses, but after usage they do return to the warehouse they belong to. The problem is to determine how many service tools of each type are needed to meet the service levels agreed upon with the customers.

In order to optimize the stock levels of service tools, first an evaluation model is needed. In this paper, therefore, we study an evaluation model for a single-location, multi-item inventory system for service tools. Different demand streams occur following a Poisson process, where for each demand stream a given set of tools is requested. When a demand occurs, all available tools are sent to the customer, and the other tools are sent from a warehouse in another region or from a central warehouse that serves as a backup. For the latter shipments, the fastest available transport mode is usually used to avoid long down-times of machines. For the warehouse under consideration, the demand for these tools may be considered as lost. Tools that are sent to the customer, return to their original location after a deterministic return time. We evaluate the *order fill rate*, i.e., the percentage of orders for which all requested tools are delivered from stock.

Currently, the evaluation of the order fill rates for service tools is done by ignoring the coupling between the demands of different tools (and thus the positive correlation between inventory levels for different items). This way of analyzing may lead to a significant underestimation of the service offered to customers (see Section 4). Our objective is to find an accurate and efficient evaluation procedure for the order fill rates. We aim at a procedure that is sufficiently efficient to be used in an optimization procedure for multi-item service tools models (in which case many evaluations have to be executed). For these optimization procedures, we may think of similar procedures as developed for multi-item spare parts models; see e.g. Wong et al. (2005) and Kranenburg and Van Houtum (2007).

In spare parts research this usage of building blocks to come up with more sophisticated models can also be recognized. Already in 1963, Hadley and Whitin (1963) studied a model very similar to ours, namely an $(S-1,S)$ policy for spare parts with a Poisson arrival process, arbitrary supply lead time distributions and lost sales. However, a difference between spare parts and service tools is that service tools are often demanded in combination with other service tools, i.e., we have *coupled demands*, while spare parts fail one at a time in general. The first model by Hadley and Whitin (1963) has been used a lot in other spare parts inventory models; see Kennedy et al. (2002), Sherbrooke (2004), and Muckstadt (2005) for an overview of the developments in this field.

The service tools problem is also related to assemble-to-order systems. In those systems several subassemblies are demanded and all have to be available before an order can be assembled. Song and Zipkin (2003) give an overview of research on assemble-to-order systems. In most of the studies backlogging is assumed, but there are also a few papers where the lost sales case is considered. Song et al. (1999) study a generalized model that has both complete backlogging and lost sales as a special case. In addition, they distinguish total order service, which means that an order is fulfilled completely or rejected as a whole, and partial order service, which means that partial fulfilment occurs as in our service tools problem. Song et al. (1999) derive an exact matrix-analytic solution for the order-fulfillment performance measures. The supply system in this paper is modeled as a single-machine exponential production facility per item. Iravani et al. (2003) extended this work by introducing flexible customers, i.e., customers that are willing to compromise on the requested items. Dayanik et al. (2003) study computationally efficient performance estimates for the same problem. When comparing our model to these assemble-to-order models, we observe the same structure for demand streams, and the return times in our model are like the lead times in an assemble-to-order system. In the terminology of assemble-to-order systems, the supply system in our model is modeled as an ample server system with equal deterministic service times for all tools. I.e., tools demanded together will return together after an equal deterministic return time for all tools, or, in other words, we have *coupled returns*. In essence, it is because of these coupled returns that the type of solutions for the assemble-to-order systems described above does not work for our problem.

Another problem related to ours is the repair kit problem; see e.g. Brumelle and Granot (1993), Mamer and Smith (1982, 1985), and Mamer and Shogan (1987). In this problem, repairmen travel around to repair machines with a repair kit containing several items. One or more items are needed to repair a machine. Thus, for a repair a subset of tools or spare parts is needed, as in our problem. The problem is to determine the optimal set of items to include in the repair kit. However, in most literature studying the repair kit problem, each repair is studied independently, which means that all tools are restocked again directly after usage, while we study demand over an infinite horizon, including the effect that tools are not readily available after usage because of the return times. More recently, Teunter (2006) studied the problem in which a repairman visits multiple locations before his repair kit is restocked. However, in this work every tour is considered separately, which means that also in this paper lead times are not considered.

Also related is the work of Güllü and Köksalan (2007). They study an optimization model for the kit-management problem with an exact evaluation. In this problem items, for instance hospital implants, are stocked at a central location, and if needed kits are composed from these items and sent to a customer's site. From the kit one item is used, and the others are returned to the central location after a certain holding time. The item that is used is replenished through a finite capacity queue. If an item is not on stock, it is supplied exogenously through an emergency channel. As soon as a unit of that item becomes available again at the central location, it is returned to the exogenous source. There are two differences between this model and the one we study in this paper. The first difference is that in our problem all service tools are returned together, while in this paper the item that is used is replenished separately after a replenishment time. The second difference is that in our problem the exact tool that is borrowed from another location is returned to that location after usage, and not the first unit of the same type that becomes available.

The main contribution of this paper is as follows. First, we introduce a new problem, in which both coupled demands and coupled returns occur. Second, we show that the full evaluation problem decomposes into evaluation problems for small sets of service tools. Third, we show that the steady-state behavior in these smaller subproblems is almost 100% insensitive to whether the distribution of the return times is deterministic or exponential. Fourth, because of this insensitivity property, we formulate three approximate Markovian models in which exponential return times are assumed. The first approximate model leads to an underestimation in all instances tested, while the second approximate model leads to an overestimation for all instances. Both approximate procedures are accurate for specific instances and can be computed efficiently. The third approximate model combines the other two, and leads to very accurate approximations for all instances tested. The absolute inaccuracy for the order fill rates as found in our test bed is only 0.005. Fifth, in the numerical experiment, we incorporate the currently used evaluation method which ignores the coupling in demands for different tools. The currently used method underestimates order fill rates by no less than 0.070 in our test bed. Hence, the third approximate model improves the currently used method by more than a factor 10.

This paper is organized as follows. In Section 2, we present our model. Section 3 shows the insensitivity of this model when the distribution of the return times is changed from a deterministic to an exponential distribution. In Section 4, the approximations are described, and numerical results are presented to show their accuracy. Finally, in Section 5, conclusions are drawn.

2. Model description

In this section, we first present our model. Afterwards, the decomposition of this model into subproblems is described in Subsection 2.1. Because of this decomposition, we only need to study small subproblems. In Subsection 2.2, we first analyze the fill rates for individual tools, and by that subproblems including only one tool are solved. Subproblems including multiple tools will be the focus of the remainder of this paper.

We consider a single warehouse with a stock of service tools, in a region where multiple machines are installed. When a machine breaks down, a demand occurs for service tools that are needed to repair the machine. Let $I = \{1, 2, ..., |I|\}$ denote the set of service tools. For each type of defect, a different subset of service tools is needed. In other words, there are different demand streams for different sets of tools. Let $K = \{1, 2, ..., |K|\}$ denote the set of demand streams, where $1 \leq |K| \leq 2^{|I|} - 1$. For each $k \in K$ the subset of service tools demanded is given by $I_k \subset I$ ($I_k \neq \emptyset$). We assume that for all demand streams at most one tool of type $i \in I$ is needed.

If all service tools included in a demanded set are available at the warehouse, the whole set is sent to the customer immediately. However, if only part of the demanded set is available, partial fulfillment of the order is possible. Thus we consider a partial order service model (cf. the terminology of Song et al., 1999). This means that the available service tools are sent to the customer immediately, and the rest of the set is sent by emergency supply (i.e., from a warehouse in another region or from a central warehouse that serves as a backup). For the stock at the warehouse, the demand for the non-available tools is considered as lost. After usage all tools are returned to their original location after a deterministic return time t_{ret}. So, there is a fixed circulation stock for each tool stocked at the warehouse under consideration, and tools always are either on stock or at a machine for repair purposes. In the latter case, a tool is said to be in the *return pipeline*. The inventory policy for the service tools in the warehouse is called a base stock policy with base stock vector $S = (S_1, S_2, \ldots, S_I)$, where S_i is the base stock level for service tool i.

The demand process for each stream $k \in K$ is a Poisson process with a constant rate λ_k (≥ 0). The demand process for all demand streams together then also is a Poisson process, with rate $\lambda = \sum_{k \in K} \lambda_k$. The assumption of Poisson demand streams follows from the technical nature of the machines under consideration. For these systems the mean time between failures is (close-to) exponential. Besides that, the warehouse serves many machines that all fail independently with low failure rates. The total demand process for this situation can be approximated well by a Poisson process. Furthermore, in practice, only short down-times of machines are allowed. Therefore, it is reasonable to assume constant failure rates.

The performance of the system is measured in terms of *order fill rates*, i.e. the percentages of orders for which all requested tools can be delivered from stock. Define β_k as the order fill rate for demand stream $k \in K$. The performance of the whole system is measured by the *aggregate order* *fill rate* β, which denotes the fraction of all demands that is fulfilled from stock:

$$\beta \quad = \quad \sum_{k \in K} \frac{\lambda_k}{\lambda} \beta_k.$$

To determine the aggregate order fill rate, we need to calculate the order fill rates for all demand streams $k \in K$. Therefore, our objective is to determine these order fill rates at given base stock levels.

2.1 Decomposition into subproblems

In real life situations, the number of service tools $|I|$ can be very large, which seems to complicate the evaluation of the order fill rates in the model described. However, because of the partial order fulfillment we can decompose our problem into smaller subproblems. Only demand for service tools that are not available is lost for the warehouse under consideration. The other tools are delivered from stock. This means that the out of stock situation for one tool does not influence the stock levels of the other service tools. Stock levels of tools are therefore only influenced by the aggregate demand for the tool itself. This also means that when determining the order fill rate for a demand stream $k \in K$, you only have to consider the subset of tools $I_k \subset I$. Hence, our problem decomposes into smaller subproblems.

Let us define *subproblem* k as the problem in which we study the order fill rate for demand stream $k \in K$. The set of tools for this subproblem is denoted by $I_k \subset I$, and the base stock levels are given by S_i for all $i \in I_k$. All tools are returned to their original location after an equal deterministic return time t_{ret}. For each $J \subset I_k$ a demand stream exists with rate $\lambda_J = \sum_{l \in K_J} \lambda_l$, where $K_J := \{l \in K | I_k \supset J \text{ and } I_l \cap (I_k \backslash J) = \emptyset\}$. This means that we consider all possible subsets of tools within the subproblem k, including the empty set, and define the aggregate demand for a specific subset as the sum over all demand streams in the original problem that do include this specific subset, but no other tools within I_k. Properties following from this definition are:

1. $\{K_J\}_{J \subset I_k}$ is a partition of K. This means that each demand stream $k \in K$ is taken into account exactly once when studying subproblem k.

2. $\sum_{J \subset I_k} \lambda_J = \lambda$. This follows from property 1.

3. $\lambda_\emptyset$ is the demand that does not include any tools within the set I_k and therefore can be neglected when determining the order fill rate for demand stream k. So, although the total demand rate is equal to the demand rate in the original problem, part of the demand is not asking for any tool within the set I_k and thus does not influence the performance of demand stream k.

For later usage, we define p_J as the probability that an arbitrary demand in subproblem k is for demand stream J ($\neq \emptyset$); i.e., $p_J := \frac{\lambda_J}{\lambda - \lambda_\emptyset}$. Finally, for each subproblem k, we define L_k as the number of demand streams $J \neq \emptyset$ for which $\lambda_J > 0$.

2.2 Fill rates for individual tools

In this subsection we study the fill rates for individual tools. We consider a service tool $i \in I$, and define $\tilde{\lambda}_i = \sum_{k \in K; i \in I_k} \lambda_k$ as the aggregate demand rate for service tool i, and $\tilde{\beta}_i$ as the fill rate of the individual service tool i. For later usage, we define $\rho_i = \frac{\tilde{\lambda}_i t_{\text{ret}}}{S_i}$ as the utilization rate of tool i, and $\tilde{p}_i$ as the probability that an arbitrary demand asks (among others) service tool i; i.e., $\tilde{p}_i = \frac{\tilde{\lambda}_i}{\lambda}$. For a service tool i, the steady-state behavior of the on-hand stock is independent of the behavior for all other tools. It is easily seen that its behavior is identical to the behavior of the number of customers in an $M/G/c/c$ queueing system with arrival rate $\tilde{\lambda}_i$, mean service time t_{ret}, and S_i servers. The fill rate then is equal to one minus the corresponding Erlang loss probability:

$$\tilde{\beta}_i \quad = \quad 1 - \frac{(\tilde{\lambda}_i t_{ret})^{S_i}/S_i!}{\sum_{j=0}^{S_i} (\tilde{\lambda}_i t_{ret})^j / j!}. \tag{1}$$

The Erlang loss probability is known to be insensitive to the distribution of the service times (see e.g. Cohen, 1976), and thus the fill rates for individual tools in our model are insensitive to the distribution of the return times.

Combining this result for individual tools and the decomposition result of the previous section, we immediately get a solution for subproblems concerning only one tool. Namely, the order fill rate for a subproblem k with $I_k = \{i\}$ for some $i \in I$ is equal to $\tilde{\beta}_i$. Hence, for the remainder of this paper, it suffices to focus on subproblems k in which more that one service tool is demanded, i.e., with $|I_k| > 1$.

3. An insensitivity result for the return times

Deterministic return times complicate the analysis of our model. I.e., having exponentially distributed return times would allow us to use Markov processes to derive the performance. Therefore, in this section, we study whether our model is sensitive to changing the deterministic return times into exponentially distributed return times. This gives a first approximate model. For subproblems with two or more tools, we will establish an almost 100% insensitivity for the order fill rate, and that insight will be used later on for the development of more efficient approximate models. Notice that, in Subsection 2.2, we established a complete insensitivity for the (order) fill rate in subproblems with only one tool.

The first approximate model, M_0, is obtained from the original model by replacing the deterministic return times by exponentially distributed return times. In this model, we keep the property that service tools demanded together are also returned together. Obviously, the behavior of M_0 may be described by a Markov process. In the state description we have to incorporate which units in the return pipeline are coupled. As orders may be filled partially, a coupling may occur for all combinations of tools. An appropriate state description would be to denote for each subset $I' \subset I_k$ how many groups of precisely the tools of I' are present in the return pipeline. This would lead

Table 1: *Parameter settings*

Name of parameter	**Number of values**	**Values**
Number of tools $\|I_k\|$ and number of demand streams L_k	5	$\|I_k\| = 2,\ L_k = 3$
		$\|I_k\| = 3,\ L_k = 4$
		$\|I_k\| = 3,\ L_k = 7$
		$\|I_k\| = 5,\ L_k = 6$
		$\|I_k\| = 3,\ L_k = 4$, asymmetric
Service level	2	low, high
Coupling factor F_k	3	0.2, 0.5, 0.8
Aggregate demand rate $\tilde{\lambda}_i$	3	0.2, 0.6, 1.0

to a $(2^{|I_k|} - 1)$-dimensional state space. For the resulting Markov process, except for very small instances, no analytical solution is available and a numerical solution will lead to large computation times because of the large size of the state space. Summarizing, the coupled returns complicate the state description and lead to a large state space, which makes the determination of the order fill rates via the Markov process unattractive. Furthermore, we are interested in insights on the sensitivity of the model to motivate the approach used for the approximate models later on in the paper. To gain these insights, for both the original model and the approximate model M_0, we use simulation to determine the order fill rates and their differences.

We now define the test bed of subproblems. W.l.o.g., for each subproblem, we may assume that $\lambda_\emptyset = 0$ and that the service tools are numbered $1, \ldots, |I_k|$ (and thus $I_k = \{1, \ldots, |I_k|\}$). The input parameters for each subproblem are the number of service tools $|I_k|$, the number of non-empty demand streams L_k, the total demand rate λ, the return time t_{ret}, the probabilities $p_J, J \subset I_k$, and the base stock levels S_i, $i \in I_k$. These input parameters are varied such that specific values are obtained for the following variables: (i) the combination of the number of tools $|I_k|$ and the number of demand streams L_k; (ii) the service level, which is determined by the choice for the basestock levels; (iii) the so-called coupling factor that is defined below; (iv) the aggregate demand rates $\tilde{\lambda}_i$. The settings for these parameters are given in Table 1, and their combinations result into $5 \cdot 2 \cdot 3 \cdot 3 = 90$ instances in total.

As can be seen in Table 1, for the combination of the number of tools and the number of demand streams five values are chosen. First, we vary the number of service tools, i.e., we take $|I_k| = 2, 3, 5$. Then, for the demand streams, we take one stream with coupled demands for the whole set of tools and streams for individual tools; $L_k = |I_k| + 1$ in that case. For $|I_k| = 3$, we also study the cases in which all possible demand streams occur and thus $L_k = 7$. We mainly keep instances symmetric, but for $|I_k| = 3$ and $L_k = 4$ we also consider asymmetric instances. Within symmetric instances, the demand rates for each service tool are equal, and also the demand rates for all subsets with the same size are equal.

Within all instances, equal base stock levels are chosen, i.e., $S_1 = \ldots = S_{|I_k|}$. For these base stock levels, two values are chosen. The first value, a base stock level of 1, leads to a low order fill

Table 2: *Parameter settings - demand probabilities*

Symmetric instances	F_k	p_J $\lvert J\rvert=1$	$\lvert J\rvert=2$	$\lvert J\rvert=3$	$\lvert J\rvert=5$
$\lvert I_k\rvert=2,\ L_k=3$	0.8	1/6	4/6		
	0.5	1/3	1/3		
	0.2	4/9	1/9		
$\lvert I_k\rvert=3,\ L_k=4$	0.8	1/7		4/7	
	0.5	1/4		1/4	
	0.2	4/13		1/13	
$\lvert I_k\rvert=3,\ L_k=7$	0.8	1/13	1/13	7/13	
	0.5	4/19	1/19	4/19	
	0.2	5/17	1/51	1/17	
$\lvert I_k\rvert=5,\ L_k=6$	0.8	1/9			4/9
	0.5	1/6			1/6
	0.2	4/21			1/21
Asymmetric instances	F_k	p_J $J=\{1\}$	$J=\{2\}$	$J=\{3\}$	$J=\{1,2,3\}$
$\lvert I_k\rvert=3,\ L_k=4$, asymmetric	0.8	1/7	2/7	0	4/7
	0.5	1/4	7/20	3/20	1/4
	0.2	4/13	5/13	3/13	1/13

rate for the demand stream that asks all tools I_k. The second value is chosen such that the order fill rates for the latter demand stream are approximately 95%. This is done to study both a low and a high service level.

We define the *coupling factor* $F_{i,k}$ for a specific subproblem k and a tool $i \in I_k$ as follows:

$$F_{i,k} = \sum_{J \subset I_k; i \in J} \frac{p_J}{\tilde{p}_i} \frac{|J|-1}{|I_k|-1}.$$

This means that $F_{i,k} = 1$ if $\tilde{p}_i = p_{I_k}$, i.e., if demand for tool i comes only from the stream demanding all tools; and $F_{i,k} = 0$ if $\tilde{p}_i = p_{\{i\}}$, i.e. if all demand for tool i comes from demands for this single tool only. Next, we define the coupling factor F_k for subproblem k as a weighted average of all individual coupling factors of the service tools:

$$F_k = \sum_{i \in I_k} \frac{\tilde{p}_i}{\sum_{j \in I_k} \tilde{p}_j} F_{i,k}.$$

We vary the probabilities p_J such that we get values of 0.2, 0.5 and 0.8 for the coupling factor F_k, which corresponds to weak, medium and strong coupling. The values p_J for the different settings for the combination of $|I_k|$ and L_k are given in Table 2.

Finally, the demand rates λ are chosen as follows. In the symmetric instances, λ is set such that for all $i \in I_k$ the aggregate demand rates $\tilde{\lambda}_i$ are equal to 0.2, 0.6 and 1.0, respectively. In the asymmetric cases, the demand rate is varied such that values of 0.2, 0.6 and 1.0 are obtained for the aggregate demand rate of service tool 1 and that the aggregate demand rates for service tools 2 and 3 are 120% and 80% of the aggregate demand rate for tool 1, respectively. Notice that instances with equal aggregate demand rates for individual tools have also equal utilization rates

Table 3: *Simulation details*

Parameter	**Unit**	**Value**
Number of simulation runs		100
Warm-up period	# demands	5000
Length of each simulation run	# demands	25000

of the tools, and therefore effects of varying coupling factors under equal utilization rates can be observed.

The return time (= mean return time in approximate model M_0), t_{ret}, is equal to 1 in all instances.

The parameters are partly chosen based on real life data from the company we collaborate with. In this company, the return time is one week. Further, an aggregate demand rate of one per week per tool is high, and aggregate demand rates of 0.2 per week is low. Also the size of the subproblems, thus the amount of tools included in demand streams, is based on actual data. Namely, in most of the real-life demands for tools up to five service tools are demanded. The base stock levels are chosen in such a way that they lead to low order fill rates, and to an order fill rate of approximately 95%. This last number is used as a target for the order fill rates and thus corresponds to a high service level. Finally, the coupling factor is varied between three values to get a better understanding of the influence of this parameter on the sensitivity of the model and later on the performance of the approximations.

Both the original model and the approximate model M_0 are studied using simulation. The details of the simulation runs per instance are given in Table 3. The simulations took on average 38.04 seconds, with a maximum of 51.49 seconds.

In Tables 4 and 5, for the symmetric and the asymmetric instances, respectively, we have listed 95% confidence intervals for the order fill rates in both models and for their differences. Here β_k denotes the order fill rate of the stream that ask all tools I_k in subproblem k in the original model with deterministic return times; and $\beta_k^{(0)}$ denotes the corresponding order fill rate for the approximate model M_0 with exponential return times. (Recall that, within subproblem k we are interested in the order fill rate for the stream that asks all tools. Order fill rates for demand streams $J \subset I_k$ with $|J| < |I_k|$ can be analyzed via smaller subproblems. The results for these demand streams are therefore not included.)

The listed results show that the differences in order fill rates are very small. The average absolute difference of $\beta_k^{(0)} - \beta_k$ is 0.0006, and the maximum absolute difference is 0.0060. In most of the instances the 95% confidence interval of the difference contains the value 0, however for some instances the value 0 is not included. This means that the model is not completely insensitive to the replacement of the deterministic return times by exponentially distributed return times.

Table 4: *Results of the sensitivity analysis - Symmetric instances*

$\lvert I_k \rvert$	L_k	$\tilde{\lambda}_i$	F_k	**Service**	S_i	β_k	$\beta_k^{(0)}$	$\beta_k^{(0)} - \beta_k$
2	3	0.2	0.8	l	1	0.8006 ± 0.0006	0.8006 ± 0.0007	0.0000 ± 0.0006
				h	2	0.9780 ± 0.0003	0.9780 ± 0.0003	0.0000 ± 0.0003
			0.5	l	1	0.7570 ± 0.0010	0.7575 ± 0.0012	0.0006 ± 0.0010
				h	2	0.9723 ± 0.0004	0.9722 ± 0.0005	-0.0001 ± 0.0004
			0.2	l	1	0.7183 ± 0.0013	0.7187 ± 0.0016	0.0004 ± 0.0014
				h	2	0.9685 ± 0.0007	0.9691 ± 0.0008	0.0006 ± 0.0007
		0.6	0.8	l	1	0.5577 ± 0.0008	0.5572 ± 0.0010	-0.0005 ± 0.0009
				h	3	0.9719 ± 0.0004	0.9717 ± 0.0003	-0.0002 ± 0.0003
			0.5	l	1	0.4810 ± 0.0012	0.4808 ± 0.0015	-0.0002 ± 0.0014
				h	3	0.9653 ± 0.0005	0.9651 ± 0.0005	-0.0003 ± 0.0005
			0.2	l	1	0.4226 ± 0.0021	0.4227 ± 0.0020	0.0000 ± 0.0019
				h	3	0.9622 ± 0.0008	0.9614 ± 0.0009	-0.0009 ± 0.0009
		1	0.8	l	1	0.4162 ± 0.0008	0.4160 ± 0.0009	-0.0003 ± 0.0009
				h	4	0.9773 ± 0.0003	0.9772 ± 0.0003	-0.0001 ± 0.0003
			0.5	l	1	0.3332 ± 0.0010	0.3331 ± 0.0014	-0.0003 ± 0.0014
				h	4	0.9723 ± 0.0005	0.9723 ± 0.0005	0.0000 ± 0.0005
			0.2	l	1	0.2769 ± 0.0019	0.2783 ± 0.0019	0.0014 ± 0.0022
				h	4	0.9701 ± 0.0006	0.9698 ± 0.0006	-0.0003 ± 0.0006
3	4	0.2	0.8	l	1	0.7716 ± 0.0007	0.7721 ± 0.0008	0.0005 ± 0.0007
				h	2	0.9728 ± 0.0003	0.9726 ± 0.0003	-0.0002 ± 0.0003
			0.5	l	1	0.6908 ± 0.0012	0.6910 ± 0.0013	0.0002 ± 0.0013
				h	2	0.9613 ± 0.0006	0.9615 ± 0.0006	0.0002 ± 0.0006
			0.2	l	1	0.6189 ± 0.0024	0.6203 ± 0.0021	0.0015 ± 0.0022
				h	2	0.9549 ± 0.0012	0.9543 ± 0.0012	-0.0005 ± 0.0011
		0.6	0.8	l	1	0.5065 ± 0.0009	0.5059 ± 0.0011	-0.0006 ± 0.0010
				h	3	0.9645 ± 0.0004	0.9642 ± 0.0004	-0.0003 ± 0.0004
			0.5	l	1	0.3795 ± 0.0014	0.3787 ± 0.0016	-0.0008 ± 0.0015
				h	3	0.9518 ± 0.0007	0.9516 ± 0.0007	-0.0002 ± 0.0008
			0.2	l	1	0.2886 ± 0.0022	0.2887 ± 0.0025	0.0001 ± 0.0026
				h	3	0.9449 ± 0.0012	0.9446 ± 0.0012	-0.0004 ± 0.0014
		1	0.8	l	1	0.3604 ± 0.0009	0.3589 ± 0.0010	-0.0015 ± 0.0011
				h	4	0.9708 ± 0.0004	0.9708 ± 0.0004	0.0000 ± 0.0004
			0.5	l	1	0.2349 ± 0.0012	0.2323 ± 0.0014	-0.0028 ± 0.0014
				h	4	0.9611 ± 0.0006	0.9613 ± 0.0007	0.0002 ± 0.0008
			0.2	l	1	0.1600 ± 0.0021	0.1582 ± 0.0023	-0.0017 ± 0.0025
				h	4	0.9567 ± 0.0011	0.9557 ± 0.0011	-0.0011 ± 0.0013
3	7	0.2	0.8	l	1	0.7846 ± 0.0007	0.7851 ± 0.0008	0.0005 ± 0.0007
				h	2	0.9749 ± 0.0003	0.9746 ± 0.0003	-0.0002 ± 0.0003
			0.5	l	1	0.7019 ± 0.0014	0.7018 ± 0.0013	-0.0001 ± 0.0014
				h	2	0.9626 ± 0.0006	0.9627 ± 0.0005	0.0002 ± 0.0005
			0.2	l	1	0.6243 ± 0.0026	0.6247 ± 0.0025	-0.0004 ± 0.0025
				h	2	0.9551 ± 0.0013	0.9549 ± 0.0013	-0.0001 ± 0.0013
		0.6	0.8	l	1	0.5245 ± 0.0010	0.5249 ± 0.0012	0.0004 ± 0.0010
				h	3	0.9665 ± 0.0004	0.9664 ± 0.0004	-0.0002 ± 0.0004
			0.5	l	1	0.3919 ± 0.0015	0.3914 ± 0.0017	-0.0006 ± 0.0016
				h	3	0.9525 ± 0.0008	0.9525 ± 0.0007	0.0000 ± 0.0008
			0.2	l	1	0.2928 ± 0.0025	0.2937 ± 0.0028	0.0011 ± 0.0031
				h	3	0.9443 ± 0.0013	0.9449 ± 0.0015	0.0006 ± 0.0016
		1	0.8	l	1	0.3764 ± 0.0013	0.3766 ± 0.0010	0.0001 ± 0.0010
				h	4	0.9723 ± 0.0004	0.9723 ± 0.0004	-0.0001 ± 0.0004
			0.5	l	1	0.2437 ± 0.0014	0.2417 ± 0.0016	-0.0020 ± 0.0015
				h	4	0.9614 ± 0.0006	0.9620 ± 0.0007	0.0005 ± 0.0008
			0.2	l	1	0.1627 ± 0.0024	0.1613 ± 0.0025	-0.0012 ± 0.0026
				h	4	0.9567 ± 0.0013	0.9564 ± 0.0012	-0.0005 ± 0.0016
5	6	0.2	0.8	l	1	0.7172 ± 0.0015	0.7325 ± 0.0015	0.0000 ± 0.0014
				h	2	0.9626 ± 0.0007	0.9659 ± 0.0007	0.0002 ± 0.0006
			0.5	l	1	0.5769 ± 0.0019	0.5771 ± 0.0018	0.0001 ± 0.0019
				h	2	0.9422 ± 0.0009	0.9419 ± 0.0009	-0.0003 ± 0.0009
			0.2	l	1	0.4630 ± 0.0048	0.6114 ± 0.0053	-0.0009 ± 0.0040
				h	2	0.9287 ± 0.0022	0.9531 ± 0.0021	0.0000 ± 0.0023
		0.6	0.8	l	1	0.4215 ± 0.0017	0.4326 ± 0.0019	-0.0001 ± 0.0016
				h	3	0.9515 ± 0.0008	0.9547 ± 0.0008	-0.0009 ± 0.0009
			0.5	l	1	0.2422 ± 0.0016	0.2378 ± 0.0016	-0.0044 ± 0.0016
				h	3	0.9286 ± 0.0010	0.9285 ± 0.0011	-0.0002 ± 0.0010
			0.2	l	1	0.1422 ± 0.0042	0.2805 ± 0.0049	-0.0005 ± 0.0051
				h	3	0.9130 ± 0.0022	0.9441 ± 0.0023	0.0012 ± 0.0029
		1	0.8	l	1	0.2748 ± 0.0016	0.2773 ± 0.0017	0.0002 ± 0.0018
				h	4	0.9602 ± 0.0007	0.9637 ± 0.0007	-0.0001 ± 0.0008
			0.5	l	1	0.1244 ± 0.0013	0.1183 ± 0.0013	-0.0060 ± 0.0014
				h	4	0.9422 ± 0.0009	0.9428 ± 0.0009	0.0005 ± 0.0009
			0.2	l	1	0.0557 ± 0.0037	0.1519 ± 0.0040	0.0027 ± 0.0050
				h	4	0.9311 ± 0.0020	0.9561 ± 0.0024	-0.0003 ± 0.0028

Table 5: *Results of the sensitivity analysis - Asymmetric instances*

$\lvert I_k \rvert$	L_k	$\tilde{\lambda}_1$	F_k	**Service**	S_i	β_k	$\beta_k^{(0)}$	$\beta_k^{(0)} - \beta_k$
3	4	0.2	0.8	l	1	0.7725 ± 0.0007	0.7730 ± 0.0008	0.0005 ± 0.0007
				h	2	0.9718 ± 0.0003	0.9716 ± 0.0003	-0.0002 ± 0.0003
			0.5	l	1	0.6915 ± 0.0013	0.6917 ± 0.0013	0.0002 ± 0.0013
				h	2	0.9606 ± 0.0006	0.9606 ± 0.0006	0.0000 ± 0.0006
			0.2	l	1	0.6196 ± 0.0023	0.6209 ± 0.0021	0.0014 ± 0.0022
				h	2	0.9537 ± 0.0011	0.9532 ± 0.0012	-0.0004 ± 0.0011
		0.6	0.8	l	1	0.5092 ± 0.0009	0.5092 ± 0.0011	0.0000 ± 0.0010
				h	3	0.9615 ± 0.0004	0.9612 ± 0.0004	-0.0003 ± 0.0004
			0.5	l	1	0.3817 ± 0.0014	0.3810 ± 0.0015	-0.0007 ± 0.0014
				h	3	0.9492 ± 0.0007	0.9490 ± 0.0007	-0.0003 ± 0.0008
			0.2	l	1	0.2903 ± 0.0022	0.2905 ± 0.0024	0.0002 ± 0.0027
				h	3	0.9424 ± 0.0011	0.9422 ± 0.0012	-0.0002 ± 0.0013
		1	0.8	l	1	0.3633 ± 0.0009	0.3635 ± 0.0009	0.0002 ± 0.0010
				h	4	0.9666 ± 0.0004	0.9667 ± 0.0005	0.0000 ± 0.0005
			0.5	l	1	0.2370 ± 0.0012	0.2349 ± 0.0014	-0.0023 ± 0.0014
				h	4	0.9577 ± 0.0007	0.9579 ± 0.0007	0.0001 ± 0.0008
			0.2	l	1	0.1615 ± 0.0021	0.1606 ± 0.0022	-0.0008 ± 0.0025
				h	4	0.9534 ± 0.0011	0.9524 ± 0.0011	-0.0011 ± 0.0013

However, since the differences are small, we conclude that replacing the deterministic return times by exponentially distributed times does lead to accurate approximations for the order fill rates.

4. Efficient approximate models

By the results of Section 3, model M_0 leads to accurate approximations for the order fill rates of the original model. As stated earlier, the behavior of M_0 can be described by a Markov process. However, for this Markov process, no analytical solution is available and a numerical solution will lead to large computation times because of the large size of the state space. Therefore, we apply additional approximations to model M_0. That leads to the approximate models M_1 and M_2, which both are obtained from M_0 by aggregation of states. Furthermore, we combine approximate model M_1 and M_2 into a third approximate model, M_3. Approximate models M_1 and M_2 are described in Subsection 4.1. Then, in Subsection 4.2, approximate model M_3 is described. After that, the accuracy of all three models is tested and compared to the currently used method in Subsection 4.3.

4.1 Description of models M_1 and M_2

The approximate models M_1 and M_2 are obtained from model M_0 by aggregating all states with the same numbers of tools in the return pipeline. In other words, in M_1 and M_2 we only keep track of the total amount of tools that is in the return pipeline and we ignore the way tools were demanded. So, when the system is in a given aggregate state, it is not known whether the tools in the return pipeline were demanded individually or in sets of tools. For the latter, we may assume two extremes, leading to M_1 and M_2, respectively. In approximate model M_1, we assume that

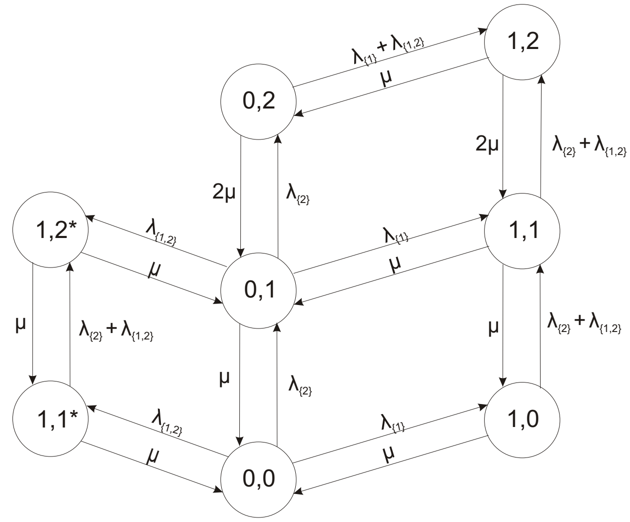

Figure 1: *Markov process for approximate model M_0 in Example 1*

all tools in the return pipeline were demanded individually and thus that they will be returned individually. I.e., in model M_1, we assume *minimal coupling for the returns*. In approximate model M_2, we assume *maximal coupling*; the precise formulation of this model follows below. We first show the construction of both M_1 and M_2 for an example.

Example 1. Consider model M_0 for a situation with $|I_k| = 2$, $L_k = 3$, $\lambda_{\{1\}} > 0$, $\lambda_{\{2\}} > 0$, $\lambda_{\{1,2\}} > 0$, and $S = (1, 2)$. The behavior of this system may be described by a Markov process. As states we define the tuples $(0, 0)$, $(0, 1)$, $(0, 2)$, $(1, 0)$, $(1, 1)$, $(1, 1)^\star$, $(1, 2)$, and $(1, 2)^\star$. For each state, the i-th component denotes the number of service tools i in the return pipeline, $i = 1, 2$. In both state $(1, 1)$ and state $(1, 1)^\star$, there is one tool of type 1 and one tool of type 2 in the return pipeline; in $(1, 1)^\star$ they are coupled and in $(1, 1)$ they are not. In both state $(1, 2)$ and state $(1, 2)^\star$, there is one tool of type 1 and two tools of type 2 in the return pipeline; in $(1, 2)^\star$ one of the two tools 2 is coupled with tool 1 and in $(1, 2)$ there is no coupling. The Markov process that is obtained via this state description is depicted in Figure 1. In this picture, $\mu = 1/t_{\text{ret}}$ denotes the return rate.

We now aggregate the states $(1, 1)$ and $(1, 1)^\star$ in an aggregated state $(1, 1)$ and the states $(1, 2)$ and $(1, 2)^\star$ in an aggregated state $(1, 2)$. For the aggregated states, we have to define the outgoing transition rates. The states that are aggregated have the same transitions that correspond to demand arrivals, and thus those transitions are taken over for the aggregated states. However, the states that are aggregated have different transitions corresponding to returns. For the aggregated states, we either assume minimal coupling, which leads to approximate model M_1 or we assume maximal coupling, which leads to approximate model M_2. The resulting Markov processes are

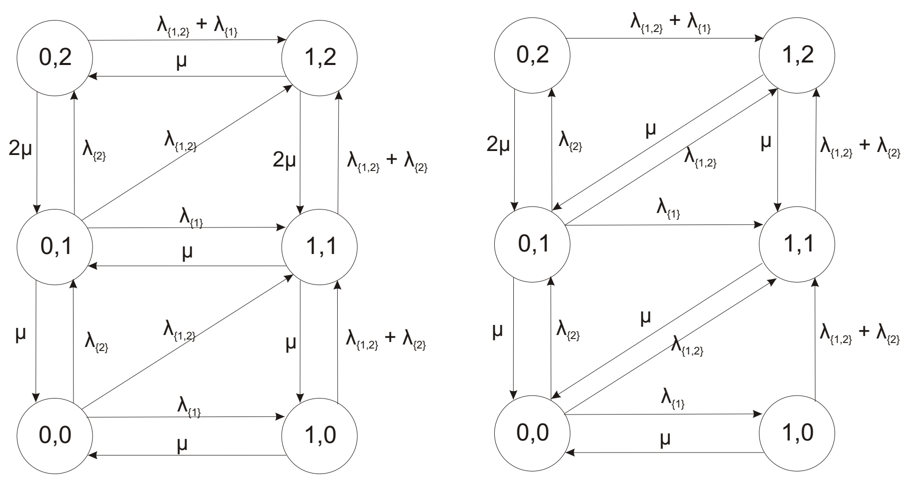

Figure 2: *Markov processes for approximate models M_1 (left) and M_2 (right) in Example 1*

denoted in Figure 2. ◊

The general description of the approximate models M_1 and M_2 is as follows. For both models for a specific subproblem k, we define aggregated states $x = (x_1, \dots, x_{|I_k|})$, where x_i denotes the number of tools i in the return pipeline, $i \in I_k$, $0 \leq x_i \leq S_i$. A state transition only occurs when a tool or a coupled group of tools is returned or when a demand occurs.

Both approximate models have the same state transitions due to demand for tools. These outgoing transitions and corresponding transition rates are as follows:

- *Transition due to demand (Model M_1 and M_2):*
  A demand for service tools $J \subset I_k$ occurs with rate λ_J. This results in a transition to state $\hat{x} = (\hat{x}_1, ..., \hat{x}_{|I_k|})$, with
$$\hat{x}_i = \begin{cases} x_i + 1 & \text{if } i \in J \text{ and } x_i < S_i; \\ x_i & \text{otherwise.} \end{cases}$$

For approximate model M_1, the outgoing transitions due to a return of a tool and corresponding transition rates from a state x are as follows:

- *Transitions due to a return (Model M_1):*
  For each $i \in I_k$ with $x_i > 0$, a tool i is returned with rate x_i/t_{ret}. This results in a transition to state $\hat{x} = (\hat{x}_1, \dots, \hat{x}_{|I_k|})$, with $\hat{x}_i = x_i - 1$ and $\hat{x}_j = x_j$ for all $j \neq i$.

For approximate model M_2, for each state x we have the same outgoing transitions due to demand as for model M_1, but for the transitions due to a return, we obtain:

- *Transitions due to a return (Model M_2):*
  We assume maximal coupling for the tools in the return pipeline. Let $G(x) = \max_{i \in I_k} x_i$ be

the maximum number of units of the same service tool in the return pipeline. We divide the tools in the return pipeline into G groups of coupled tools. Group g, $g = 1, \ldots, G$, consists of the tools i for which $x_i \geq g$. This implies that tool i is part of the first x_i groups, and tool i is not part of the other groups. Notice that it is possible that two groups consist of precisely the same tools. Each group g returns with rate $1/t_{\text{ret}}$ and this results in a transition to state $\hat{x} = (\hat{x}_1, \ldots, \hat{x}_{|I_k|})$, with

$$\hat{x}_i = \begin{cases} x_i - 1 & \text{if } x_i \geq g \\ x_i & \text{otherwise} \end{cases}$$

For both M_1 and M_2, the number of states is $\prod_{i \in I_k}(S_i + 1)$. The steady-state distribution may be computed from the steady-state equations via successive substitutions. This is efficient as long as the number of tools in a subproblem $|I_k|$ and the base stock levels S_i are small (notice that this is the case in the representative test bed of Section 3). An approximation for the order fill rate β_k of the original model is obtained by the summation of the steady-state probabilities over all states x for which $x_i < S_i$ for all $i \in I_k$. The resulting approximations obtained via M_1 and M_2 are denoted by $\beta_k^{(1)}$ and $\beta_k^{(2)}$, respectively.

Notice that when there is no coupling at all between the demands for different tools, M_1 will be identical to M_0, while when there is 100% coupling M_2 will be identical to M_0. This suggests that M_1 will lead to accurate approximations for the order fill rates in the original model for instances with low coupling factors, and probably somewhat less accurate approximations for instances with high coupling factors; similarly, M_2 will lead to accurate approximations for the order fill rates in the original model for instances with high coupling factors, and probably somewhat less accurate approximations for instances with low coupling factors.

4.2 Description of approximate model M_3

As explained in the last section, approximate model M_1 will be most accurate when the coupling factor is low, while approximate model M_2 will be most accurate for a high coupling factor. This observation leads to the third approximate model that combines the other two. Model M_3 takes a weighted average of model M_1 and M_2, where the weight factor depends on the coupling factor. The order fill rate $\beta_k^{(3)}$ obtained by approximate model M_3 is given by:

$$\beta_k^{(3)} = (1 - F_k)\beta_k^{(1)} + F_k \beta_k^{(2)}.$$

This weighted average will lead to the same results as model M_1 when $F_k = 0$, in which case model M_1 is identical to model M_0. The weighted average will lead to the same results as model M_2 when $F_k = 1$, in which case model M_2 is identical to model M_0.

4.3 Computational Results

The accuracy of the approximate models M_1, M_2 and M_3 has been tested on the basis of the test bed of Section 3. In Tables 6 and 7, for the symmetric and the asymmetric instances, respectively, we have listed the differences in the order fill rates $\beta_k^{(1)}$, $\beta_k^{(2)}$ and $\beta_k^{(3)}$ obtained via M_1, M_2 and M_3 compared to the order fill rates β_k of the original model. Also the performance of the method as used currently is added for comparison. In this method the dependency between the demands for service tools is ignored. Let us define β_k^c as the order fill rate in subproblem k for the currently used method. These order fill rates are calculated as follows:

$$\beta_k^c = \prod_{i \in I_k} \tilde{\beta}_i = \prod_{i \in I_k} (1 - \frac{(\tilde{\lambda}_i t_{ret})^{S_i}/S_i!}{\sum_{j=0}^{S_i} (\tilde{\lambda}_i t_{ret})^j/j!}).$$

The computation time per instance was on average 1.51, 1.46 and 2.97 seconds for model M_1, model M_2 and model M_3 respectively. The maximum computation times was 38.30, 36.77 and 75.04 seconds for model M_1, model M_2 and model M_3 respectively. The computations were executed on a Pentium 4 PC.

For all instances, the differences $\beta_k^{(1)} - \beta_k$ are negative and the differences $\beta_k^{(2)} - \beta_k$ are positive. Thus, M_1 leads to underestimations for the order fill rates in all instances, while M_2 leads to overestimations. This is directly related to the fact that we assume minimal and maximal coupling, respectively, for tools in the repair pipeline. Minimal coupling in returns leads to a minimal correlation between on-hand stocks of the tools, which, intuitively, has a negative effect on order fill rates, and for maximal coupling this is the other way around.

The accuracy of the approximations obtained via M_1 varies from reasonable to very good. The absolute difference $|\beta_k^{(1)} - \beta_k|$ is on average equal to 0.038. For instances with respectively weak, medium and strong coupling, the averages are 0.010, 0.034 and 0.071, and thus M_1 is indeed most accurate for low coupling factors. When the service levels increase due to higher base stock levels, the performance becomes better; the average absolute difference is 0.069 when all base stock levels are 1, which is the low service case, versus 0.007 with higher base stock levels, which lead to a high service. For different failure rates, there is no significant difference in the performance of model M_1. The averages for the absolute difference $|\beta_k^{(1)} - \beta_k|$ for low, medium and high demand rates are 0.031, 0.044 and 0.040, respectively. For symmetric and asymmetric instances, the absolute errors are 0.038 and 0.038, respectively. This means that there is no significant difference between the performance of the approximate models for the symmetric and the asymmetric instances. Summarizing, we can say that M_1 performs well when the fill rates are high, and then especially when the coupling factor is low.

The accuracy of the approximations obtained via M_2 also varies from reasonable to very good. The absolute difference $|\beta_k^{(2)} - \beta_k|$ is on average equal to 0.033. For instances with respectively weak, medium and strong coupling, the averages are 0.046, 0.035 and 0.019, and thus M_2 is indeed

Table 6: *Results for M_1, M_2, M_3, and the current method - Symmetric instances*

$\lvert I_k \rvert$	L_k	$\tilde{\lambda}_i$	F_k	**Service**	S_i	β_k	$\beta_k^c - \beta_k$	$\beta_k^{(1)} - \beta_k$	$\beta_k^{(2)} - \beta_k$	$\beta_k^{(3)} - \beta_k$
2	3	0.2	0.8	l	1	0.801	-0.106	-0.057	0.006	-0.007
				h	2	0.978	-0.011	-0.007	0.001	-0.001
			0.5	l	1	0.757	-0.063	-0.032	0.012	-0.010
				h	2	0.972	-0.005	-0.003	0.001	-0.001
			0.2	l	1	0.718	-0.024	-0.012	0.017	-0.006
				h	2	0.969	-0.001	-0.001	0.001	0.000
		0.6	0.8	l	1	0.558	-0.167	-0.098	0.024	-0.001
				h	3	0.972	-0.011	-0.008	0.001	-0.001
			0.5	l	1	0.481	-0.090	-0.050	0.045	-0.002
				h	3	0.965	-0.005	-0.003	0.002	-0.001
			0.2	l	1	0.423	-0.032	-0.017	0.058	-0.002
				h	3	0.962	-0.001	-0.001	0.002	0.000
		1	0.8	l	1	0.416	-0.166	-0.104	0.038	0.010
				h	4	0.977	-0.008	-0.006	0.001	0.000
			0.5	l	1	0.333	-0.083	-0.048	0.067	0.010
				h	4	0.972	-0.003	-0.002	0.002	0.000
			0.2	l	1	0.276	-0.027	-0.014	0.080	0.005
				h	4	0.970	-0.001	0.000	0.002	0.000
3	4	0.2	0.8	l	1	0.772	-0.193	-0.091	0.010	-0.011
				h	2	0.973	-0.021	-0.013	0.001	-0.002
			0.5	l	1	0.691	-0.112	-0.051	0.024	-0.014
				h	2	0.961	-0.010	-0.005	0.002	-0.002
			0.2	l	1	0.619	-0.040	-0.017	0.039	-0.006
				h	2	0.955	-0.003	-0.002	0.002	-0.001
		0.6	0.8	l	1	0.507	-0.262	-0.148	0.037	0.000
				h	3	0.965	-0.023	-0.015	0.002	-0.001
			0.5	l	1	0.380	-0.135	-0.072	0.075	0.002
				h	3	0.952	-0.010	-0.006	0.005	-0.001
			0.2	l	1	0.289	-0.045	-0.022	0.102	0.003
				h	3	0.945	-0.003	-0.002	0.004	-0.001
		1	0.8	l	1	0.360	-0.235	-0.147	0.056	0.016
				h	4	0.971	-0.016	-0.011	0.003	0.000
			0.5	l	1	0.235	-0.110	-0.064	0.098	0.017
				h	4	0.961	-0.007	-0.004	0.004	0.000
			0.2	l	1	0.160	-0.035	-0.019	0.118	0.008
				h	4	0.957	-0.002	-0.001	0.004	0.000
3	7	0.2	0.8	l	1	0.785	-0.206	-0.100	0.009	-0.013
				h	2	0.975	-0.023	-0.015	0.001	-0.002
			0.5	l	1	0.702	-0.123	-0.059	0.023	-0.018
				h	2	0.963	-0.011	-0.007	0.002	-0.002
			0.2	l	1	0.624	-0.046	-0.021	0.038	-0.009
				h	2	0.955	-0.004	-0.002	0.002	-0.001
		0.6	0.8	l	1	0.525	-0.280	-0.160	0.037	-0.002
				h	3	0.967	-0.025	-0.017	0.002	-0.002
			0.5	l	1	0.392	-0.148	-0.080	0.075	-0.002
				h	3	0.953	-0.011	-0.007	0.005	-0.001
			0.2	l	1	0.293	-0.049	-0.024	0.103	0.001
				h	3	0.944	-0.003	-0.001	0.005	0.000
		1	0.8	l	1	0.376	-0.251	-0.158	0.058	0.015
				h	4	0.972	-0.018	-0.013	0.002	-0.001
			0.5	l	1	0.244	-0.119	-0.069	0.101	0.016
				h	4	0.961	-0.007	-0.004	0.005	0.000
			0.2	l	1	0.163	-0.038	-0.021	0.119	0.007
				h	4	0.957	-0.002	-0.001	0.004	0.000
5	6	0.2	0.8	l	1	0.717	-0.315	-0.130	0.018	-0.011
				h	2	0.963	-0.042	-0.021	0.002	-0.003
			0.5	l	1	0.577	-0.175	-0.070	0.048	-0.011
				h	2	0.942	-0.022	-0.010	0.005	-0.003
			0.2	l	1	0.463	-0.061	-0.023	0.081	-0.002
				h	2	0.929	-0.008	-0.004	0.006	-0.002
		0.6	0.8	l	1	0.422	-0.326	-0.183	0.058	0.010
				h	3	0.952	-0.047	-0.026	0.005	-0.002
			0.5	l	1	0.242	-0.147	-0.078	0.115	0.019
				h	3	0.929	-0.024	-0.012	0.010	-0.001
			0.2	l	1	0.142	-0.047	-0.025	0.142	0.009
				h	3	0.913	-0.008	-0.004	0.011	-0.001
		1	0.8	l	1	0.275	-0.244	-0.163	0.082	0.033
				h	4	0.960	-0.035	-0.021	0.005	0.000
			0.5	l	1	0.124	-0.093	-0.059	0.126	0.034
				h	4	0.942	-0.017	-0.009	0.009	0.000
			0.2	l	1	0.056	-0.025	-0.014	0.137	0.016
				h	4	0.931	-0.006	-0.003	0.010	-0.001

Table 7: *Results for M_1, M_2, M_3, and the current method - Asymmetric instances*

$\|I\|$	L_k	λ_i	CF	**Service**	S_i	β_k	$\beta_k^c - \beta_k$	$\beta_k^{(1)} - \beta_k$	$\beta_k^{(2)} - \beta_k$	$\beta_k^{(3)} - \beta_k$
3	4	0.2	0.8	l	1	0.773	-0.193	-0.091	0.009	-0.011
				h	2	0.972	-0.021	-0.013	0.001	-0.002
			0.5	l	1	0.692	-0.112	-0.051	0.023	-0.014
				h	2	0.961	-0.010	-0.006	0.002	-0.002
			0.2	l	1	0.620	-0.040	-0.017	0.038	-0.006
				h	2	0.954	-0.003	-0.002	0.002	-0.001
		0.6	0.8	l	1	0.509	-0.264	-0.148	0.034	-0.002
				h	3	0.962	-0.022	-0.015	0.002	-0.001
			0.5	l	1	0.382	-0.136	-0.072	0.073	0.000
				h	3	0.949	-0.010	-0.006	0.004	0.000
			0.2	l	1	0.290	-0.045	-0.022	0.100	0.003
				h	3	0.942	-0.003	-0.002	0.005	0.000
		1	0.8	l	1	0.363	-0.237	-0.148	0.053	0.013
				h	4	0.967	-0.015	-0.010	0.002	0.000
			0.5	l	1	0.237	-0.111	-0.064	0.096	0.016
				h	4	0.958	-0.006	-0.004	0.004	0.000
			0.2	l	1	0.162	-0.035	-0.019	0.116	0.008
				h	4	0.953	-0.002	-0.001	0.004	0.000

most accurate when there is strong coupling. When the service levels increase due to higher base stock levels, the performance becomes better; the average absolute difference is 0.063 when all base stock levels are 1, which is the low service case, versus 0.004 with higher base stock levels, when the service is high. For different failure rates, there are differences in the performance of model M_2. The averages for the absolute difference $|\beta_k^{(2)} - \beta_k|$ for low, medium and high demand rates are 0.015, 0.038 and 0.047 respectively. Approximate model M_2 thus performs better for lower failure rates. For symmetric and asymmetric instances, the absolute errors are 0.034 and 0.032, respectively. This means that there is no significant difference between the performance of the approximate models for the symmetric and the asymmetric instances. Summarizing, we can say that M_2 performs well when the fill rates are high, and then especially when the coupling factor is high.

The accuracy of approximate model M_3 is very good. The average absolute difference between model M_3 and the original model, $|\beta_k^{(3)} - \beta_k|$, is 0.005. For instances with respectively weak, medium and strong coupling, the averages are 0.003, 0.007 and 0.006, and thus M_3 is almost equally accurate for high, medium and low coupling factors. When the service levels increase due to higher base stock levels, the performance becomes better; the average absolute difference is 0.010 when all base stock levels are 1, which situation leads to low service levels, versus 0.001 with higher base stock levels, which lead to high service levels. For different failure rates, there is no significant difference in the performance of model M_3. The averages for the absolute difference $|\beta_k^{(3)} - \beta_k|$ for low, medium and high demand rates are 0.006, 0.002 and 0.008 respectively. For symmetric and asymmetric instances, the absolute errors are 0.006 and 0.005, respectively. This means that there is no significant difference between the performance of the approximate models for the symmetric and the asymmetric instances. Summarizing, we can say that M_3 performs well in all instances. The maximum absolute error is 0.034, which is about the same as the average performance of the other approximations.

The currently used method leads to an underestimation of the service offered to customers. The average difference between this approximation and the original model, $\beta_k^c - \beta_k$ is -0.070. If we compare the average absolute differences of approximate models M_1 and M_2 with the current method, we see that both models lead to an improvement of almost 50%, which is a significant improvement. However, for approximate model M_3, we even see an improvement of over 90% compared to the current method.

In conclusion, approximate model M_3 leads to efficient and accurate approximations in all instances tested. Therefore, M_3 is definitely appropriate to be used in an optimization algorithm for the stock levels of service tools. The models M_1 and M_2 might be less appropriate, since they lead to poor approximations in some cases.

5. Conclusion

We studied an evaluation model for the order fill rate in a multi-item inventory system for service tools, with coupled demands and coupled returns. We showed that this full evaluation model decomposes into smaller subproblems. For these subproblems, we showed that the steady-state behavior is almost insensitive for whether return times are deterministic or exponential. Based on that insight, we formulated three approximate models M_1, M_2 and M_3, and we found that model M_3 leads to efficient and fairly accurate approximations in all considered instances.

The comparison with the current way of evaluating the order fill rates shows that there is a large improvement possible by incorporating the coupled demands in approximate evaluations. This encourages more research in this field. First of all, the proposed approximate evaluation model can be used in an optimization model for service tools in a single location. Later on, this can be extended to a model including multiple locations to resemble practice even more.

References

Brumelle S. and D. Granot, 1993, The repair kit problem revisited, Operations Research 41, 994-1006.

Cohen J., 1976, On regenerative processes in queueing theory, Springer, Berlin, Germany.

Dayanik S., J. Song and S. Xu, 2003, The effectiveness of several performance bounds for capacitated production, partial-order-service, assemble-to-order systems, Manufacturing & Service Operations Management 5, 230-251.

Güllü R. and M. Köksalan, 2007, A new model for the kit-management problem, working paper, Bogazici University, Istanbul, Turkey.

Hadley G., and T. Whitin, 1963, Analysis of Inventory Systems, Prentice-Hall, Englewood Cliffs, NJ, USA.

Iravani S., K. Luangkesorn and D. Simchi-Levi, 2003, On assemble-to-order systems with flexible customers, IIE Transactions 35, 389-403.

Kennedy W., J. Patterson and L. Fredendall, 2002, An overview of recent literature on spare parts inventories, International Journal of Production Economics 76, 201-215.

Kranenburg A. and G. Van Houtum, 2007, Effect of commonality on spare parts provisioning costs, International Journal of Production Economics, To appear.

Mamer J. and S. Smith, 1982, Optimizing field repair kits based on job completion rate, Management Science 28, 1328-1333.

Mamer J. and S. Smith, 1985, Job completion based inventory systems: Optimal policies for repair kits and spare machines, Management Science 31, 703-718.

Mamer J. and A. Shogan, 1987, A constrained capital budgeting problem with applications to repair kit selection, Management Science 33, 800-806.

Muckstadt J., 2005, Analysis and Algorithms for Service Parts Supply Chains, Springr, Berlin, Germany.

Sherbrooke C., 2004, Optimal Inventory Modeling of Systems: Multi-Echelon Techniques, Kluwer, Dordrecht, The Netherlands.

Song J., S. Xu and B. Liu, 1999, Order-fulfillment performance measures in an assemble-to-order system with stochastic leadtimes, Operations Research 47, 121-149.

Song J. and P. Zipkin, 2003, Supply chain operations: Assemble-to-order systems, in: T. de Kok and S. Graves, eds., Handbooks in Operations Research and Management Science, Vol. 11, Supply Chain Management: Design, Coordination and Operation, Elsevier, Amsterdam, The Netherlands.

Teunter R., 2006, The multi-job repair kit problem, European Journal of Operational Research 175, 1103-1116.

Wong H., G. Van Houtum, D. Cattrysse, and D. Van Oudheusden, 2005, Simple, efficient heuristics for multi-item, multi-location spare parts systems with lateral transshipments and waiting time constraints, Journal of the Operational Research Society 56, 1419-1430.