# Stock keeping unit fill rate specification

**[Accepted for publication by the *European Journal of Operational Research*]**


**R.H. Teunter[1], A.A. Syntetos[2 #,] and M.Z. Babai[3]**

[1] University of Groningen, NL, [2] Panalpina Centre for Manufacturing and Logistics Research, Cardiff University, UK, [3] Kedge Business School, FR

[#] Corresponding author: +44 (0) 2920 876572, SyntetosA@cardiff.ac.uk



# ABSTRACT

The fill rate is the most widely applied service level measure in industry and yet there is minimal advice available on how it should be differentiated on an individual Stock Keeping Unit (SKU) basis given that there is an overall system target service level. The typical approach utilized in practice, and suggested in academic textbooks, is to set the individual service levels equal to the targeted performance required across an entire stock base or a certain class of SKUs (e.g., in ABC classification). In this paper it is argued that this approach is far from optimal and a simple methodology is proposed that is shown (on real life datasets) to be associated with reductions in stock investments. In addition, the new approach is intuitive, very easy to implement and thus highly likely to be positively received by practitioners and software manufacturers.



**Keywords:** Inventory; Service Level; Fill Rate; Safety Stock.


# 1. INTRODUCTION

For inventory control, service levels constitute arguably the most important performance measures (Silver et al., 1998). The fill rate in particular (which determines the percentage of demand satisfied directly from stock-on-hand) is the most commonly used measure in industry, as it translates directly to the customer service level achieved (e.g., Guijarro et al., 2012). This measure, also known as the volume fill rate, is to be distinguished from the order fill rate which represents the fraction of complete orders that can be filled directly from inventory (Larsen and Thorstenson, 2008, 2014). Another, somewhat less common, service measure is the ready rate, specified as the fraction of time during which the stock-on-hand is positive. It is well known (e.g., Silver et al., 1998, Axsäter, 2006) that for pure Poisson and normally distributed demand the (volume) fill rate is equivalent (although only approximately in the case of the normality assumption) to the ready rate. Other service level measures used in inventory systems are reviewed by Schneider (1981) and Silver and Bischak (2011).

Service level targets drive the determination of safety stocks and thus inventory investments and the responsiveness of the system to market (step) changes. Such targets should relate explicitly to individual Stock Keeping Units (SKUs). From an analytical perspective, safety stock and ordering calculations are performed at the individual item level. It is also intuitively appealing that different items must receive different treatment based on their characteristics. However, the considerable number of SKUs that modern organisations deal with often implies that targets are being set at the aggregate / *system* level, be that the entire stock base or a category of SKUs following a particular classification approach (e.g., ABC classification) (Teunter et al., 2010). The target service level then assigned to individual SKUs is simply that at the system level. One would expect that this approach is far from optimal when contrasting it to looking at each item separately and yet there are no guidelines in the literature as to how the latter could be done to differentiate service levels for individual SKUs in order to meet an overall system target service level. The specification of the 'right' service level on an individual SKU basis constitutes the purpose of this research.

## 1.1. Practical and Research Background

The authors have encountered a great number of companies that use the ABC classification to set service levels, by assigning the same service level to each SKU in a particular class. This is in line with findings from Lee (2002) from NONSTOP solutions (a provider of demand-chain optimization services) and Pflitsch (2008) from SLIMSTOCK (a provider of forecasting and inventory management software, including *Slimstock ABC* for inventory classification). Both confirm from their extensive experience of implementing inventory control software that the standard approach is to fix service levels per class.

Consideration of the individual SKU requirements and the assignment of a service level that is thought to be appropriate for a specific SKU occurs only on an exceptional basis. Companies are known for example to do so for what are often termed the '*super A*' items (in an ABC type classification), i.e. items of exceptional contribution to the profit margin or revenues.

State-of-the-art inventory planning software does offer manual (trial-and-error) experimentation to trade-off service levels and cycle/safety stock on an individual SKU basis. Inventory examination capabilities of that nature are supported by most specialized inventory control solutions, for example, *RightStock* from DBO SERVICES (Dawson, 2013) as well as supply chain planning software packages, for example, the *Inventory Planner* of ARKIEVA (Singh, 2013). In Figure 1 we depict the *Inventory Strategist* function of *RightStock.*

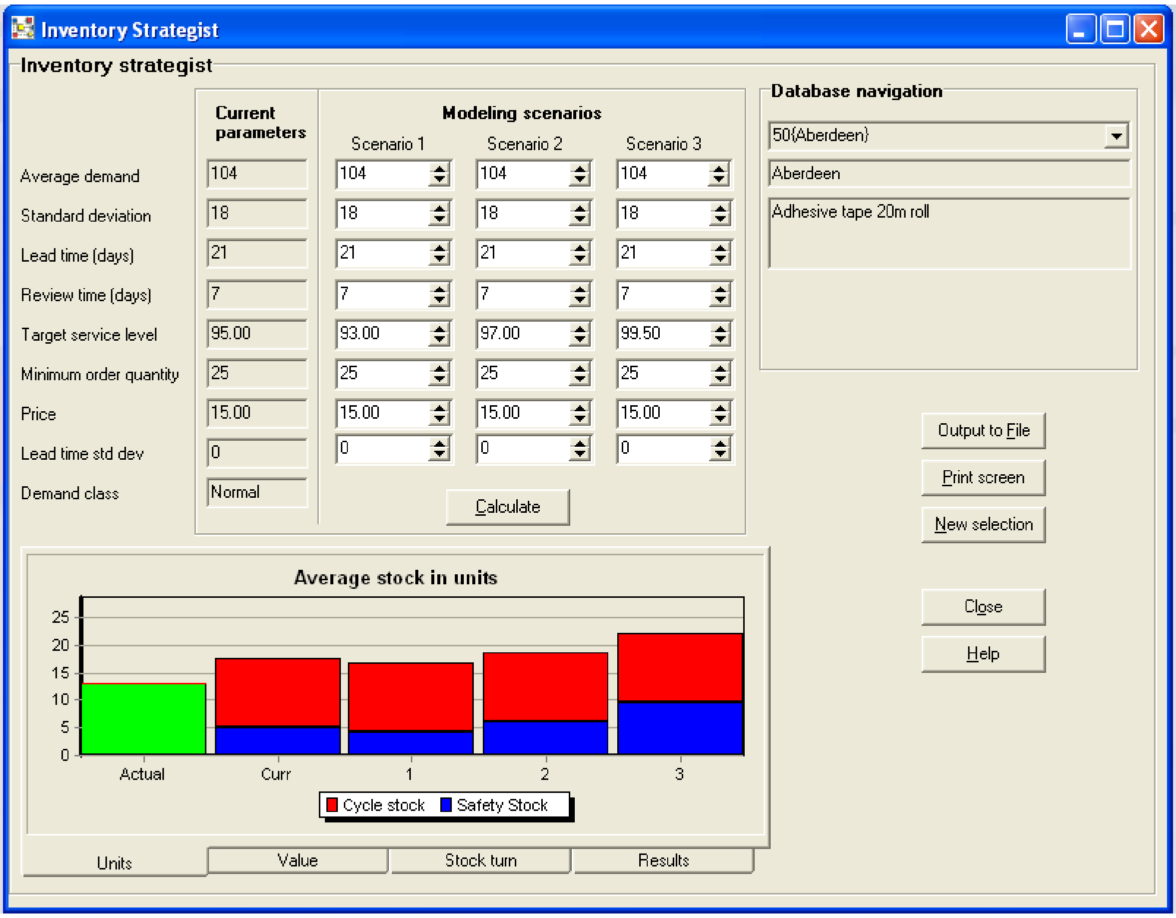


Figure 1. Inventory Strategist screenshot from RightStock software
(presented with permission granted from DBO Services Limited)

For the example considered in Figure 1, the system target service level is set to 95% and managers may experiment with the implications of altering the target for a particular SKU in terms of the average inventory required to sustain such a target. Inventory is controlled by a continuous re-order point, order quantity policy. This results in the cycle inventory being half of the order quantity; the average inventory equals the cycle inventory plus the safety stock (determined in this example based on the normality assumption). The implications of altering the service target, ceteris paribus, are automatically recalculated and managers may experiment to reach a decision as to what is the ‘best’ service level for a specific SKU.

Obviously, manual exploration of the effect of service level variation on the system cost and the system service level can only be done for a small selection of SKUs. Furthermore, lack of insight into why one SKU gets a higher service level than another may prevent managers from following such advice. Making such evaluations across the entire range of SKUs requires a system optimization approach that, to the best of our knowledge, is not available in any commercial inventory planning tool.

Although the majority of the inventory literature is also concerned with single SKU systems, the literature *does* suggest some multi-SKU approaches for setting service levels. Most of these approaches are rather complex and may therefore be difficult to implement. We refer interested readers to Thonemann et al. (2002) for a discussion of key findings, and will only discuss the relatively simple approaches here.

Motivated by the absence of simple multi-SKU approaches, despite their great practical relevance, Thonemann et al. (2002) set out to derive an "easy-to-use model to estimate the benefit of using a system approach". Their model is restricted to a spare parts environment with Poisson demand and base stock ordering policies (with order quantities equal to one), where all parts have the same lead time. Moreover, their model and method are arguably still complex to apply in the real world as they involve "continuous approximations and a quadratic fit".

Hopp et al. (1997) also note that practical system approaches are lacking in the literature, and present three heuristics for finding near optimal reorder points and order quantities across a range of SKUs. As Thonemann et al. (2002), they consider a spare parts environment with Poisson demand. They allow batch demands but of a constant size, which can also be interpreted as assuming pure Poisson demand for the relevant demand unit. All three heuristics determine policy parameters using separable closed-form expressions, although these expressions are rather complex and difficult to update. The authors argue that "as long as this computation can be done outside the system and updated infrequently, we consider it to be within the spirit of easily implementable".

In this paper we develop a much simpler methodology for the specification of target service levels on an individual SKU basis, given an aggregate target at a group level. Furthermore, our method is applicable in any setting (where a reorder point, order quantity strategy is used) and so not restricted to spare part environments. Experimentation with three real life datasets shows cost reductions in stock investments when comparing the performance of the new approach to the standard practice of assigning to an item the service target allocated to the group and situations with service level variation across SKU classes (in ABC type classifications). The new approach is very easy to interpret and implement, and thus highly likely to be positively received by practitioners and software manufacturers. In addition, the transparency and clear algorithmic steps associated with its implementation make it a valuable tool in terms of further constructive judgemental interventions.

The remainder of the paper is organized as follows: in the next Section a method is presented for setting service levels on an individual SKU basis. In Section 3, an empirical investigation is undertaken to contrast the performance of the new method to the standard practice of setting individual SKU fill rate targets equal to that of a whole system and to two theory informed approaches based on ABC classification. The empirical data is presented, followed by a discussion of the experimental setting and the results along with their interpretation. The conclusions of this work and its implications for practice are finally considered in Section 4.

# 2. SETTING SERVICE LEVELS ON AN INDIVIDUAL SKU BASIS

We consider a multi-item inventory system where the demand process can be of any type (although we will consider normally distributed demand in our numerical investigation). The system is controlled according to the commonly studied and widely applied continuous reorder point ($r$), order quantity ($Q$) policy, from now on referred to as the ($r,Q$) policy. It is well-known that the order quantity trades off fixed ordering/setup cost and holding cost, whereas the reorder point determines the safety stock and thereby affects the service level (or, conversely, it is set up to the required amount to ensure the achievement of a pre-specified service level target). This is true if the joint optimization of the order quantity and the reorder point is not taken into account, which is the case considered in our paper. Our focus is on service levels and consequently we will treat the order quantities of the different SKUs as given. For this reason, we do not include fixed ordering costs in our analysis, since the order frequencies are not affected by reorder points and corresponding service levels. However, we remark that we shall consider the costs related to cycle stocks ($Q/2$) in our empirical investigation in Section 3, in order to be able to determine the reduction in the total costs resulting from our approach.

In this section we propose a method to set the target service levels for each SKU in the inventory system in a way that enables the achievement of a target system service level at minimum cost. Service performance is captured through the fill rate. We remark that although we consider a service level constraint rather than a penalty cost for backorders, we do introduce virtual penalty costs per time unit for each backorder. These virtual penalty costs will prove useful for our analysis, but it is important to note that it is not part of the proposed method for setting service levels that result from it (which would be very impractical).

There can be various reasons why penalty costs may differ across SKUs. In service logistics, some parts may be more critical to the functioning of equipment than others. In other industries, retailers may consider part of their assortment more critical than the rest, for instance because they are often demanded by important (large) customers. In our experience with several firms, however, quantifying relative criticality is very hard. Managers find it difficult to assess if one SKU is, say, two, five or ten times as critical as another. Moreover, as discussed before, most firms stock thousands of different SKUs and so estimating criticality figured for each one separately would be very time consuming. More realistic and manageable, in our view and experience, is to divide the complete assortment in several groups (e.g. high, medium and low criticality) and set three different group target service levels. The method that we propose can then be applied to each group separately.

If all SKUs (of the whole assortment or a selected group) are assumed to be equally critical, then the virtual penalty cost is set to the same value for all SKUs. This is in line with targeting a minimum system fill rate across all SKUs, since that system fill rate is equally affected by backorders related to any item. As we argued, in many practical situations this will be much easier to implement as it avoids having to estimate the (relative) criticality for each separate SKU. In our empirical investigation on multiple real life datasets we will also follow this approach, since criticality figures are not available. However, there may of course be situations where such figures *are* available. In order to (i) be able to deal with such situations as well, and (ii) obtain further insights into the effect of criticality on desired service levels, we do include SKU specific criticality figured into our analysis.

## 2.1. Notation

We introduce the following notation for the remainder of the paper.

$N$ : Number of SKUs in the inventory system or in a class of items

$D_i$ : Expected demand (per time unit) of SKU $i$

$FR_i$ : Fill rate of SKU $i$

$FR_T$ : System (Total) fill rate across all SKUs, i.e. $FR_T = \sum_{i=1}^{N} FR_i D_i \Big/ \sum_{i=1}^{N} D_i$

$p_i$ : Unit cost (price) of SKU $i$

$h_i$ : Unit inventory holding cost of SKU $i$ per time unit

$\alpha$ : Inventory holding charge (assumed constant for all items in a stock base), i.e. $h_i = \alpha \ p_i$

$c_i$ : (Relative) Criticality of a backlog for SKU $i$ per time unit

$b$ : Standardized (virtual) backlog cost per time unit

$b_i$ : Backlog cost $b_i = b\, c_i$ of SKU $i$ per time unit

$PCR_i$ : Price Criticality Ratio $p_i / c_i$ of SKU $i$

$APCR$ : Average Price Criticality Ratio, $APCR = \sum_{i=1}^{N} \frac{p_i}{c_i} D_i \Big/ \sum_{i=1}^{N} D_i$

$AP$ : Average Price, $AP = \sum_{i=1}^{N} p_i D_i \Big/ \sum_{i=1}^{N} D_i$

$RP$ : Relative Price of SKU $i$, $RP = p_i / AP$

$Q_i$ : Order quantity of SKU $i$

$r_i$ : Reorder point of SKU $i$

$L_i$ : (Constant) Lead-time of SKU $i$

We remark that e.g. Hopp et al. (1997) and Thonemann et al. (2002) use the same definition of the system fill rate. Note that fast moving items have the highest effect on the system fill rate, given its (average) demand weighted definition.

## 2.2. Derivation of Optimal Service Levels

For each SKU of type $i$, inventory is controlled using the reorder point, order quantity ($r_i$,$Q_i$) policy. This policy places a replenishment order of size $Q_i$ if the inventory position (stock on hand plus on order minus backorders) drops to or below $r_i$. By selecting the value of the reorder point (given the order quantity) for SKU $i$, its service level is controlled. Recall that the objective is to minimize the system (i.e. for all SKUs) inventory cost under the restriction that some minimum target system fill rate is achieved. So, there is a service level restriction at the system level rather than at the individual SKU level.

A well-known result from the inventory control theory (Axsäter, 2006, p.105) is that the optimal fill rate $FR_i$ for SKU $i$ under a reorder point, order quantity inventory control policy is given by the following newsboy formula:

$$1 - FR_i = \frac{h_i}{b_i + h_i}$$

Please note that the above formula only holds when the order quantity is (exogenously) given. As previously discussed, this is the case considered in the paper, i.e. we do not address a joint optimization of the reorder point and order quantity. Please also note that although both the holding and backorder costs are time-related, the optimality condition specifies a certain fill rate based on the ratio of these costs. This does not imply that the fill rate is the most suitable service measure for the situation with time-dependent holding and backorder costs. Indeed, its use would imply that the length of a backorder does not matter, which would provide a strong theoretical incentive never to fulfil any backordered demand. In most practical situations, this is unwanted as the length of a backorder does matter. In our analysis, we therefore build on the optimality condition given above, without implying that the fill rate (at the single SKU level) is the only relevant service level measure.

As backorders are typically unwanted, the (virtual) backorder cost $b_i$ is usually assumed to be considerably higher than the holding cost $h_i$. Therefore, we approximately get:

$$1 - FR_i \approx \frac{h_i}{b_i} = \frac{\alpha \ p_i}{b \ c_i} \tag{1}$$

We remark that we will only use this *approximation* to derive a method for varying fill rates across SKUs of a group, but will ensure that the *exact* system fill rate always achieves its target. This will become clear at the end of the section and illustrated for a specific example.

From the definition of the total fill rate $FR_T$, we directly get

$$1 - FR_T = \frac{\sum_{i=1}^{N} (1 - FR_i) D_i}{\sum_{i=1}^{N} D_i} \tag{2}$$

Using (1) and recalling that $APCR$ denotes the Average Price Criticality Ratio weighted by demand rate, this can be rewritten as

$$1 - FR_T = \frac{\alpha}{b} \frac{\sum_{i=1}^{N} \frac{p_i}{c_i} D_i}{\sum_{i=1}^{N} D_i} = \frac{\alpha}{b} APCR$$

and so we get

$$\frac{\alpha}{b} = \frac{(1 - FR_T)}{APCR} \tag{3}$$

Combining (1) and (3) gives

$$1 - FR_i \approx (1 - FR_T)\frac{PCR_i}{APCR} \tag{4}$$

Interpreting the fraction on the right hand side of (4) as the relative Price Criticality Ratio, we observe a very simple and insightful relation: a lower relative Price Criticality Ratio implies a higher desired fill rate. We obtain an even simpler interpretation for the special case where all SKUs are equally critical, i.e. where the objective is to minimize the average number of backordered demands over time. It is easy to see that the relative Price Criticality Ratio is then equal to the relative price of an SKU, and so a lower price implies a higher desired fill rate. Though intuitive, the very issue of whether low or high value items should attract more stock (and thus higher availability) has been debated extensively in the literature (see, e.g., Teunter et al., 2010).

Another important observation is that the fill rate expression (4) does not involve the standardized (virtual) backorder cost. This is an important advantage as that cost may be very difficult to obtain or even estimate in real life situation. Indeed, this explains why companies typically prefer to specify target system fill rates instead of estimating backorder costs.

We further remark that, though relative, criticality figures may also be difficult to estimate in practice. In that case, one option is of course to use the same criticality for all items and, as explained above, base fill rates through (4) on the relative prices rather than the relative price criticality ratio's. Indeed, this is what we will do in our numerical investigation for three real life datasets in Section 3, where criticality figures are not available.

It should also be noted that the fill rate value $FR_i$ resulting from (4) can be negative for very expensive SKUs. This is a consequence of the approximation underlying (1). Although $1 - h_i/(h_i + b_i)$ is always positive, $1 - h_i/b_i$ is negative if $h_i > b_i$, which may happen for exceptionally expensive SKUs if the target system fill rate is not so high and, accordingly, the (virtual) backorder cost rate is small. A simple adjustment in such a case is to set any negative fill rate values to zero. We realize that zero or very low fill rates may not be desirable in practice, even for very expensive SKUs. For such cases, our method can easily be adapted by imposing some positive lower bound on the fill rate for each individual SKU, although this does imply that the achieved system fill rate may exceed its target. We will demonstrate this empirically in Section 3 and reflect on it in Section 4.

We end this section by illustrating the use of (4) for a scenario (for demonstration purposes) with three SKUs of equal criticality, a target system fill rate of 96%, and with cost prices and demands as given in Table 1. That table also shows the calculation of the fill rates for all SKUs using (4).

Table 1. Example of the application of the new method: system fill rate target = 0.96 (96%), 3 SKUs

| $SKU_i$ | Price ( $p_i$ ) | Demand ( $D_i$ ) | $FR_i$ |
|---|---|---|---|
| 1 | 1 | 70 | 1 – (1 – 0.96) x (1 / 4) = 0.99 |
| 2 | 5 | 20 | 1 – (1 – 0.96) x (5 / 4) = 0.95 |
| 3 | 23 | 10 | 1 – (1 – 0.96) x (23 / 4) = 0.77 |

Average Price ( $AP$ ) = 0.7 x 1 + 0.2 x 5 + 0.1 x 23 = 4

Validation: 0.7 x 0.99 + 0.2 x 0.95 + 0.1 x 0.77 = 0.96

# 3. EMPIRICAL INVESTIGATION

## 3.1. Empirical Data

For the purpose of the empirical investigation, we use three datasets. Dataset 1 consists of 4,799 SKUs and comes from a warehouse supplying spare parts globally for the installed base of machines that are used in the textile industry. Dataset 2 consists of 39,274 SKUs and comes from a retailer that sells bike and car parts and accessories. Dataset 3 consists of 9,086 SKUs and comes from a retailer located in the Netherlands that sells do-it-yourself products. The datasets contain details on the demand, lead-times, order quantities and costs. Tables 2, 3 and 4 show some key descriptive statistics for datasets 1, 2 and 3 respectively. Please note that the statistics presented in these Tables are calculated across SKUs. Although all datasets appear to be quite skewed in terms of mean demand, this does not necessarily apply to individual SKU demands. We remark that the same datasets have previously been used by Teunter et al. (2010).

Table 2. Descriptive statistics of dataset 1

| 4,799 SKUs | Demand (per week) | Lead-time (weeks) | Order Quantity | St. dev. of Demand | Cost (Euro) |
|---|---|---|---|---|---|
| **Min** | 0.01 | 0.03 | 1 | 0.05 | 0.01 |
| **25%ile** | 0.28 | 0.26 | 10 | 0.55 | 0.64 |
| **Median** | 1.26 | 0.26 | 35 | 1.83 | 5.03 |
| **Mean** | 13.93 | 0.37 | 268 | 14.83 | 35.56 |
| **75%ile** | 5.76 | 0.46 | 140 | 6.55 | 28.43 |
| **Max** | 4142.20 | 2.07 | 85070 | 6400.75 | 3021.12 |

Table 3. Descriptive statistics of dataset 2

| 39,274 SKUs | Demand (per week) | Lead-time (weeks) | Order Quantity | St. dev. of Demand | Cost (Euro) |
|---|---|---|---|---|---|
| **Min** | 0.01 | 0.3 | 1 | 0.05 | 0.01 |
| **25%ile** | 0.04 | 0.3 | 1 | 0.10 | 1.73 |
| **Median** | 0.12 | 0.3 | 4 | 0.20 | 4.55 |
| **Mean** | 0.36 | 0.3 | 8 | 0.38 | 25.42 |
| **75%ile** | 0.32 | 0.3 | 10 | 0.38 | 15.00 |
| **Max** | 35.04 | 0.3 | 209 | 52.18 | 1528.44 |

Table 4. Descriptive statistics of dataset 3

| 9,086 SKUs | Demand (per week) | Lead-time (weeks) | Order Quantity | St. dev. of Demand | Cost (Euro) |
|---|---|---|---|---|---|
| **Min** | 0 | 2 | 1 | 0 | 0.01 |
| **25%ile** | 0.04 | 2 | 30 | 0.17 | 0.91 |
| **Median** | 0.13 | 2 | 60 | 0.76 | 7.02 |
| **Mean** | 2.35 | 4 | 712 | 3.24 | 13.75 |
| **75%ile** | 0.48 | 2 | 250 | 3.00 | 20.50 |
| **Max** | 8.42 | 17 | 1000000 | 15845.31 | 339.00 |

The three datasets described above indicate a wide range of SKUs varying from very slow moving ones to fast movers and from inexpensive items to very expensive ones. The SKUs are also characterized by lead-times and ordering quantities that exhibit considerable variation.

Therefore, the datasets enable collectively a thorough investigation of the performance of the approaches considered in this paper under a wide range of empirical conditions.

We remark that all three companies from which the datasets are obtained apply continuous review ($r$,$Q$) inventory policies.

## 3.2. Experimental Setting and Performance Measurement

Although the method discussed in this paper applies to any type of demand distribution, in our numerical investigation we consider Normal demand. We remark that the normal distribution may theoretically not be the best choice for all SKUs, especially the slow moving ones. However, research by Porras and Dekker (2008) shows that it often performs well empirically for inventory cost minimization, even for slow moving SKUs. See also Syntetos and Boylan (2008). Moreover, in practice, there often is limited data available to fit demand distributions and this also applies to our three datasets, as also discussed in Teunter et al. (2009). Furthermore, using the normal distribution allows a direct comparison with other benchmark methods that will be discussed in this section.

In order to evaluate the performance of our proposed method, we calculate the (expected average) system stock value resulting from the determination of the individual fill rates in the way discussed in Section 2.2. We compare that with the standard method where the fill rates are all set equal to the target system fill rate, but also with two other relevant approaches discussed in the literature (Teunter et al, 2010; Zhang et al, 2001). Both approaches have been developed in the context of ABC-type SKU classification and they are concerned with the achievement of a target system fill rate (although the service levels set per class are Cycle Service levels (1 – probability of stock-out) rather than fill rates). Unlike typical ABC classifications that distinguish between SKUs based on criteria that bear little relevance to cost optimality (most commonly the demand volume or demand value / annual dollar volume), these approaches suggest alternative inventory theory informed criteria to rank the SKUs. The criteria are as follows:

$$\text{Zhang et al. (2001) criterion: } \frac{D_i}{h_i^{\,2} L_i}$$

$$\text{Teunter et al. (2010) criterion: } \frac{b_i D_i}{h_i Q_i}$$

For each of those approaches, SKUs are ranked based on the relevant criterion and ABC class sizes are determined using the commonly applied rule that classes A, B and C contain 20%, 30% and 50% of all SKUs. Cycle Service Levels (CSL) per class are established (and reorder levels per SKU within the class specified) such that the average demand weighted fill rate across all SKUs is equal to the target system fill rate. For more details regarding the implementation of those approaches and the experimental structure outlined above please refer to Teunter et al. (2010).

In the *appendix*, we show how to calculate the system stock value. For the purpose of the empirical investigation, three target system fill rates are considered, namely: $FR_T$ = 95%, 97%, 99%. In our experience, this is the range in which most companies operate.

## 3.3. Results

We report in Table 5, for each considered value of the target system fill rate and each dataset, the system stock value resulting from using our proposed method, the standard method (where SKU fill rates are set equal to the overall target) and those proposed by Zhang et al. and Teunter et al., as well as the percentage system stock value reduction resulting from our method as compared to these alternatives. We also report the achieved fill rate resulting from our method and, for completeness, the CSLs specified per class when using the Zhang et al. and Teunter et al. approaches.

Recall from the previous section that the new method achieves a fill rate above the target if (4) proposes negative fill rates for the most expensive SKUs, which are then corrected to zero (or even higher imposed lower bounds as will also be considered in this section).

Table 5. Empirical results - minimum fill rate equal to 0%

| | **New method** | | **Standard method** | **Overall** $FR_T$ | **Class** | **Zhang et al.** | | **Teunter et al.** | | **% reduction of system stock value of new method compared to** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Achieved** $FR_T$ | **System stock value (M€)** | **System stock value (M€)** | | | ***CSL*** | **System stock value (M€)** | ***CSL*** | **System stock value (M€)** | **Standard** | **Zhang et al.** | **Teunter et al.** |
| **Dataset 1** | 96.04% | 0.88 | 2.96 | 95% | A | 0.886 | 2.02 | 0.895 | 2.01 | 70% | 56% | 56% |
| | | | | | B | 0.500 | | 0.500 | | | | |
| | | | | | C | 0.500 | | 0.500 | | | | |
| | 97.36% | 1.18 | 3.33 | 97% | A | 0.977 | 2.11 | 0.980 | 2.09 | 65% | 44% | 44% |
| | | | | | B | 0.590 | | 0.603 | | | | |
| | | | | | C | 0.500 | | 0.500 | | | | |
| | 99.01% | 2.01 | 4.03 | 99% | A | 0.987 | 2.53 | 0.990 | 2.45 | 50% | 21% | 18% |
| | | | | | B | 0.967 | | 0.969 | | | | |
| | | | | | C | 0.500 | | 0.500 | | | | |
| **Dataset 2** | 95.58% | 0.47 | 0.91 | 95% | | | | | | 48% | | |
| | 97.11% | 0.56 | 0.98 | 97% | | | | | | 43% | | |
| | 99.00% | 0.77 | 1.11 | 99% | A | 0.924 | 0.92 | 0.932 | 0.92 | 30% | 16% | 16% |
| | | | | | B | 0.768 | | 0.711 | | | | |
| | | | | | C | 0.500 | | 0.500 | | | | |
| **Dataset 3** | 95.29% | 2.42 | 4.95 | 95% | A | 0.798 | 4.39 | 0.799 | 4.35 | 51% | 45% | 44% |
| | | | | | B | 0.500 | | 0.500 | | | | |
| | | | | | C | 0.500 | | 0.500 | | | | |
| | 97.04% | 3.11 | 5.44 | 97% | A | 0.899 | 4.58 | 0.901 | 4.51 | 43% | 32% | 31% |
| | | | | | B | 0.500 | | 0.500 | | | | |
| | | | | | C | 0.500 | | 0.500 | | | | |
| | 99.00% | 4.62 | 6.32 | 99% | A | 0.976 | 5.20 | 0.985 | 5.14 | 27% | 11% | 10% |
| | | | | | B | 0.744 | | 0.679 | | | | |
| | | | | | C | 0.500 | | 0.500 | | | | |

We observe from Table 5 that the performance gap between the methods of Teunter et al. and Zhang et al. is smaller than reported in Teunter et al. (2010). The reason for this is that we focus on the relative reduction in the total inventory cost and not just the safety stock (costs). The results in Table 5 also demonstrate stock value reductions resulting from the adoption of the new approach compared to both other approaches. Reductions of the stock value as compared to the standard method are at least 27% for all datasets and considered targets and may be as high as 70%. Similarly, the stock value reductions as compared to the Zhang et al. and Teunter et al. approaches can be as high as 56%. Note that the system stock value in dataset 2 is considerably lower than that in datasets 1 and 3, which is expected since dataset 2 is associated with much lower demand. Please also note that for dataset 2 only the results for a 99% system target fill rate are presented. The reason for this is that due to the small lead time (of 2 days) for all SKUs (see Table 3), a 95% and even a 97% fill rate is achieved with zero safety stock (i.e., a cycle service level of 50%) for all SKUs.

Furthermore, the large savings associated with the new method are realized despite the fact that it achieves a considerably higher system fill rate (above the target) in most cases. Recall that this outcome results from the fact that negative fill rates (resulting from (4)) are not allowed. In these cases, even higher savings can be achieved by lowering the fill rate target used by the new method until the achieved fill rate is (approximately equal to) the (minimum) system fill rate. Using Figure 2, it is easy to pick a suitable target. Finding the one that hits the exact target may require considerable trial-and-error. However, this is more of a theoretical issue, as in practice providing managers with different combinations of achieved system fill rates and corresponding average stock values is useful anyway, as it aids them in selecting the appropriate levels.

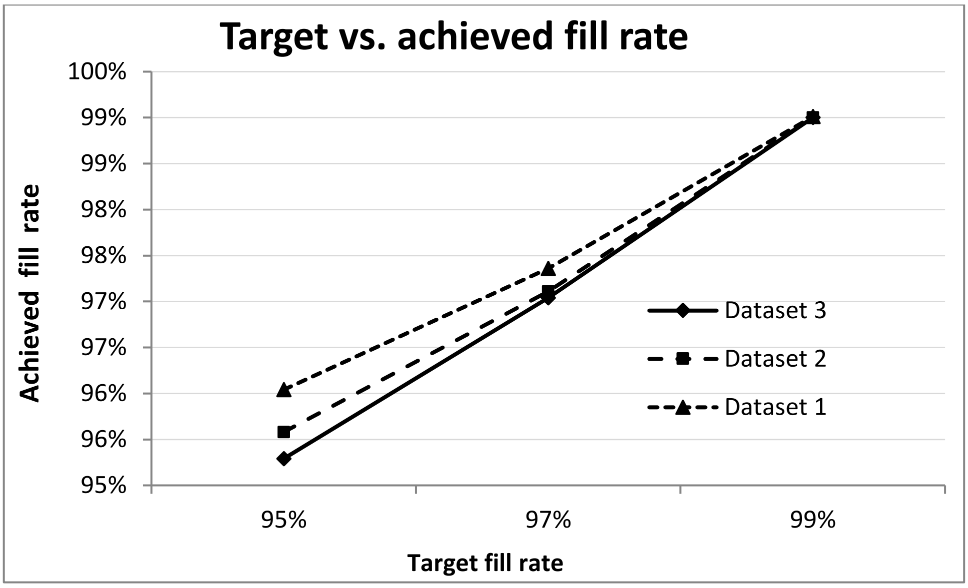


Figure 2. Target vs. achieved fill rate (minimum fill rate equal to zero)

The new method achieves its cost savings by varying the fill rate across SKUs, with higher fill rates for less expensive SKUs. In order to show the fill rate variation, we plot in Figure 3 the cumulative percentage of SKUs, for dataset 1, against the individual SKU fill rates assigned by the new method. (Results for the other datasets lead to similar insights and thus are not further discussed here.)

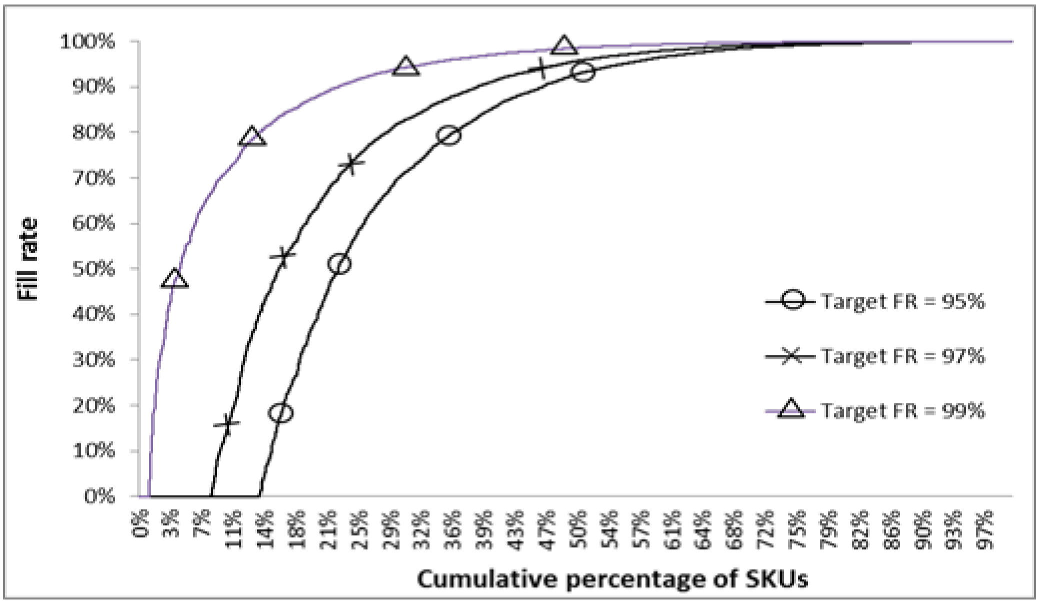


Figure 3. Distribution of the SKUs with respect to their individual fill rate (dataset 1)

Obviously, an increase in the target system fill rate leads to higher fill rates over the whole range of SKUs and also fewer SKUs with a zero fill rate. It can also be seen from Figure 3 that for targets of 95% and even 97%, a considerable portion of the SKUs have a fill rate of less than 50% under the new method. As discussed in Section 2, such low fill rates may not be desirable in all real-life settings, even for very expensive SKUs. In Table 6, we therefore report the percentage reduction of the system stock value resulting from the new method when compared to the standard one, if we impose a lower bound of 50%, 70% and 90% on the fill rate per SKU.

Table 6. Empirical results for different minimum fill rate values

| | $FR_T$ | **% reduction of system stock value compared to standard method** | | | |
|---|---|---|---|---|---|
| | | **Min FR = 0%** | **Min FR = 50%** | **Min FR = 70%** | **Min FR = 90%** |
| **Dataset 1** | **95%** | 70% | 57% | 44% | 15% |
| | **97%** | 65% | 56% | 46% | 23% |
| | **99%** | 50% | 48% | 44% | 31% |
| **Dataset 2** | **95%** | 48% | 38% | 27% | 7% |
| | **97%** | 43% | 37% | 29% | 12% |
| | **99%** | 30% | 29% | 27% | 18% |
| **Dataset 3** | **95%** | 51% | 43% | 33% | 10% |
| | **97%** | 43% | 39% | 33% | 15% |
| | **99%** | 27% | 27% | 26% | 20% |

Imposing a minimum fill rate of 50% clearly leads to an increase of the achieved system fill rate as well as an increase in the stock value. However, a comparison of Tables 5 and 6 shows that the effect is relatively small and in particular that the cost reduction is still above 27% for all datasets and all considered targets, despite the fact that the achieved service levels can be considerably above their targets. The results when imposing a minimum fill rate of 70% or 90% show that for all datasets and considered service level targets, the cost reduction offered by the new method is at least 7% (achieved for dataset 2 and $FR_T$ = 95%).

# 4. CONCLUSION AND IMPLICATIONS

A simple methodology was proposed for the specification of the target fill rate at an individual SKU level. The method works in the following steps:

1. Determine an appropriate target fill rate for a class or group of items;
2. If relative criticality figures are available, calculate the average price criticality ratio across the entire class or group of SKUs, and the price criticality ratio for each SKU. Else, calculate the average price across the entire class or group of SKUs and the relative price (see Section 2.1 for definitions);
3. Calculate the target fill rate per SKU using (4).

After calculating the target fill rates in Step 3, the inventory policies should be selected and, in particular, the order levels should be determined. This is not a specific part of our approach and therefore we do not include it as a fourth step of the proposed method.

The new method explicitly allows for variation of service levels across SKUs. An empirical investigation for three real-life datasets showed that the new method leads to cost reductions of at least 27% compared to using the same service level for each SKU, whilst overall achieved service levels are above the target. Moreover, considerable cost reductions can be achieved even if we impose a lower bound (e.g., 50%, 70% or 90%) on the fill rate per SKU to avoid very low fill rates for some SKUs, which may not be desirable in practice. Considerable cost reductions were also reported when the new method was compared to situations with service level variation across (three) classes of SKUs. Savings of at least 10% and 11% can be achieved in comparison with the approaches proposed by Teunter et al. (2010) and Zhang et al. (2001), respectively. The savings in both cases can be up to 56%.

It should be noted that although the average fill rate resulting from our method is above the target, the fraction of revenue delivered from stock-on-hand may be (much) smaller. Increasing the minimum target fill rate, as suggested above, is one practical way of dealing with this issue. Regardless, an additional benefit of the new method (although not explicitly considered in our analysis) is that it may lead to reduced obsolescence costs by stocking fewer expensive and often slow moving SKUs. This is based on the assumption that the $(r,Q)$ policy is used across all items in a stock base (or class). Although there are real world scenarios where this is the case (e.g., Syntetos et al., 2010), the employment of this policy in a slow moving demand context is not very common, at least not without a suitable adjustment of relevant parameter values (e.g., considerably higher inventory holding charges to reflect the risk of obsolescence).

There are limitations of our approach. Recall from Section 2.2 that it is based on an approximation that is accurate if target fill rates are high. Although there usually is a high system target fill rate in practice, typically in the range of 95% to 99%, our results have also shown that it is optimal to use much smaller fill rates for the most expensive (and least critical) items. So, there is no guarantee that the proposed method leads to a near-optimal solution, although the results have been very encouraging in that there were large cost savings compared to existing methods. The approximation is crucial, however, for obtaining a very simple method and in particular a closed-form and insightful relation (4) between SKU fill rates and the system fill rate. The simplicity should aid practical implementation. In fact, the method has already been incorporated by a large inventory software supplier and the simplicity was an important reason for doing so. Nevertheless, further research could explore (more) exact approaches.

Another limitation is that we did not consider possible heterogeneity amongst customers. If customers have different demand patterns, then our method may lead to poor service (below the average system fill rate across all customers) for customers who often demand expensive, slow moving SKUs, which may be undesirable. A 'solution' may be to define clusters of SKUs, but this implies a complex SKU clustering based on customer demand segmentation. Instead, an easier adjustment is to introduce minimum service levels per SKU (higher than 50%), although this does obviously imply (without further modification of the method) that the achieved system fill rate will be above the target fill rate.

Besides SKU fill rates, and the corresponding system average, order fill rates (the percentage of orders, satisfied in complete, from stock on hand) are relevant if customers regularly order multiple SKUs (e.g., Song, 1998; Syntetos et al., 2009). In such situations, if customers typically order a similar set of SKUs and delivery as a whole is essential, the collective availability across sets of SKUs matters.

Finally, our analysis assumes that the order quantity for a particular SKU is given (exogenously chosen) and thus is not addressing the interactions between setting the re-order point and order quantity.

Despite these limitations (opportunities for extending this work), our results clearly show that there is much to be gained by service level differentiation across SKUs. Moreover, we have provided a very simple and intuitive method that can serve as a starting point for both practical implementations and further research.

# ACKNOWLEDGEMENTS

We would like to thank the anonymous referees for their constructive comments that greatly helped to improve the content of our paper and its presentation.

# REFERENCES


Axsäter, S. (2006). *Inventory control*. 2nd ed. New York, NY: Springer-Verlag.

Dawson, P. (2013). Managing Director of DBO Services (http://www.dboservices.com/). Private communications to the authors.

Guijarro, E., Cardós, M., & Babiloni, E. (2012). On the exact calculation of the fill rate under discrete demand patterns. *European Journal of Operational Research, 218* (2), 442-444.

Hopp, W.J., Spearman, M.L., & Zhang, R.Q. (1997). Easily implementable inventory control policies. *Operations Research, 45* (3), 327-340.

Larsen, C., & Thorstenson, A. (2008). A comparison between the order and the volume fill rate for a base stock inventory control system under a compound renewal demand process. *Journal of the Operational Research Society, 59* (6), 798-804.

Larsen, C., & Thorstenson, A. (2014). The order and volume fill rates in inventory control systems. *International Journal of Production Economics, 147* (Part A), 13-19.

Lee, C.B. (2002). Demand Chain optimization: pitfalls and key principles, NONSTOP's "supply chain management seminar", White Paper, Retrieved from http://www.idii.com/wp/nonstop_dco.pdf

Pflitsch, R. (2008). Vice President of Slimstock B.V. (http://www.slimstock.com/).Private communication to the authors.

Porras, E., & Dekker, R. (2008.) An inventory control system for spare parts at a refinery: an empirical comparison of different re-order point methods. *European Journal of Operational Research*, *184* (1), 101-132.

Schneider, H. (1981). Effect of service-levels on order points or order levels in inventory models. *International Journal of Production Research*, *19* (6), 615-631.

Silver, E.A., & Bischak, D.P. (2011). The exact fill rate in a periodic review base stock system under normally distributed demand. *OMEGA: The International Journal of Management Science*, *39* (3), 346-349.

Silver, E.A., Pyke, D.F., & Peterson, R. (1998). *Inventory management and production planning and scheduling*, 3rd ed. New York, NY: John Wiley & Sons, Inc.

Singh, S.K. (2013). COO of Arkieva (http://www.arkieva.com). Private communications to the authors.

Song, J.S. (1998). On the order fill rate in a multi-item, base-stock inventory system. *Operations Research*, *46* (6), 831-845.

Syntetos, A.A., Babai, M.Z., Davies, J., & Stephenson, D. (2010). Forecasting and stock control: a study in a wholesaling context. *International Journal of Production Economics*, *127* (1), 103-111.

Syntetos, A.A., & Boylan, J.E. (2008). Demand forecasting adjustments for service level achievement. *IMA Journal of Management Mathematics*, *19* (2), 175-192.

Syntetos, A.A., Keyes, M., & Babai, M.Z. (2009). Demand categorisation in a European spare parts logistics network. *International Journal of Operations and Production Management, 29* (3), 292-316.

Teunter, R.H., Babai, M.Z., & Syntetos, A.A. (2010). ABC classification: service levels and inventory costs. *Production and Operations Management, 19* (3), 343-352.

Thonemann, U.W., Brown, A.O., & Hausman, W.H. (2002). Easy quantification of improved spare parts inventory policies. *Management Science*, *48* (9), 1213-1225.

Zhang, R.Q., Hopp, W.J., & Supatgiat, C. (2001). Spreadsheet implementable inventory control for a distribution centre. *Journal of Heuristics*, *7* (2), 185-203.

# APPENDIX: CALCULATION OF THE SYSTEM STOCK VALUE UNDER NORMAL DEMAND

We first introduce the following notations in addition to those in Section 2.1.

$\sigma_i$ : Standard deviation of the demand (per time unit) of SKU $i$

$n_i$ : expected backordered quantity of SKU $i$ just before a replenishment

$S_i$ : Average inventory holdingof SKU $i$

$\Phi(.)$ : Standard normal probability distribution function

$\phi(.)$ : Standard normal probability density function

As demand per time unit of SKU $i$ is i.i.d Normal with mean $D_i$ and standard deviation $\sigma_i$, lead time demand is also Normal and its mean and standard deviation are equal to $L_i D_i$ and $\sqrt{L_i}\sigma_i$, respectively. Axsäter (2006) shows that for this lead time demand distribution and under the $(r_i, Q_i)$ policy, the expected number of units of each replenishment order that arrive late (i.e. are used to satisfy backorders) is

$$n_i = \sigma_i \sqrt{L_i} \begin{bmatrix} \phi\left(\frac{r_i - L_i D_i}{\sigma_i \sqrt{L_i}}\right) - \left(\frac{r_i - L_i D_i}{\sigma_i \sqrt{L_i}}\right)\left[1 - \Phi\left(\frac{r_i - L_i D_i}{\sigma_i \sqrt{L_i}}\right)\right] \\ -\phi\left(\frac{r_i + Q_i - L_i D_i}{\sigma_i \sqrt{L_i}}\right) + \left(\frac{r_i + Q_i - L_i D_i}{\sigma_i \sqrt{L_i}}\right)\left[1 - \Phi\left(\frac{r_i + Q_i - L_i D_i}{\sigma_i \sqrt{L_i}}\right)\right] \end{bmatrix} \tag{5}$$

From this, we can calculate the fill rate of SKU $i$ as

$$FR_i = 1 - \frac{n_i}{Q_i} \tag{6}$$

Note that equations (5) and (6) above correspond to equations (5.55) and (5.54), respectively, in Axsäter (2006, p. 100).

Given $FR_i$ and $Q_i$, the reorder point $r_i$ can be easily determined from (5) and (6) by using for example Solver or Goal Seek in an Excel application.

The corresponding average stock on hand level of SKU $i$ is calculated as (Axsäter, 2006, p. 104, equation (5.65))

$$S_i = \left( r_i + \frac{Q_i}{2} - L_i D_i \right) + \frac{L_i (\sigma_i)^2}{Q} \left[ H\left( \frac{r_i - L_i D_i}{\sigma_i \sqrt{L_i}} \right) - H\left( \frac{r_i + Q_i - L_i D_i}{\sigma_i \sqrt{L_i}} \right) \right] \tag{7}$$

where

$$H(x) = \frac{1}{2}\left[ (x^2 + 1)(1 - \Phi(x)) - x\phi(x) \right]$$

The system stock value is given by

$$C_T = \sum_{i=1}^{N} p_i \, S_i$$