Józef OKULEWICZ*

VERIFICATION OF A SERVICE LEVEL ESTIMATION METHOD

Abstract: Fluctuations of supply and demand require that supplementary stock be maintained which would support satisfying the demand even with temporarily delayed supplies. It has become a common practice that what is known as the safety stock serves as a reserve until the next supply arrives. However, the importance of this deviation source has been exaggerated enormously and thus it has to be verified.The paper presents a statistical model which has served as a basis for a formula for computing the safety stock, as well as formulae for computing service level, in the two ways used. Then a simulation model follows supporting the determination of customer service level. It has been proved that the service level is a better tool for estimating the safety stock protecting against random fluctuations of supply and demand.

Keywords: safety stock, order point, service level, simulation.

1. Introduction

Where supplies to a warehouse or shop are executed at fixed time intervals and goods are collected in fixed quantities and at fixed time intervals too, there supply-demand synchronisation enables the entire demand to be satisfied. Only when fluctuations of supply and demand have occurred does the problem arise of guaranteeing supplementary supply which would support satisfying the demand even with temporarily delayed supplies. This is what is known as "safety stock". It has become a common practice that safety stock serves as a reserve protecting against random irregularities in supply or production.

While it is assumed that "inventory is working capital which costs real money" (Bonelli 2009), safety stock is not seen to be source of savings. It is just known it has to be maintained as it covers a certain proportion of supply and demand fluctuations, which contributes to service level.

Two main categories of irregularities may be identified which speak for maintaining safety stock. One of them is connected with changes in supply or production conditions. This requires supply service to be prepared appropriately, so as to be able

* Warsaw University of Technology, Warsaw, Poland

to respond to such changes promptly and effectively, and adapt the supply system to new conditions. Before this has been done, safety stock enables goods (e.g., required for production) to be collected. The response of Nokia's supply service may serve an example here. After a fire prevented the then main supplier from performing contracted supplies, Nokia's supply service managed to arrange for supplies from an alternative source in just two weeks. In the same situation, another manufacturer had to withdraw from the mobile telephone market, because its supply service failed to respond to a seemingly minor fire at the supplier (LE 2004).

The other category is connected with random deviations from agreed average values of the volumes of both supply and collection, as well as times thereof. Research into this issue resulted in the development of appropriate theory based on statistical analysis and supporting the computation of safety stock. The intended purpose of the theory was to ensure an appropriate service level meant as the probability of securing the stock for a customer. However, the importance of this deviation source has been exaggerated enormously.

Moreover, the very idea of the service level so understood raises fundamental doubts because of its definition. While assuming that the supplier-customer relation is defined in terms of "service level", it was admitted that certain customers will not be serviced. This is a kind of deficit distribution, which contradicts the very nature of logistics. The mission of logistics is to create conditions ensuring successful supplies, and not to be satisfied with failed supplies. But before the idea of service level is replaced by any idea better harmonised with the philosophy of logistics, the value of safety stock, as it is proposed in textbooks, should be verified.

2. Safety stock

Safety stock is necessary, where supplies are repeatedly delivered to a warehouse in random time intervals and customers collect random quantities of the goods supplied. Assuming that the warehouse stock is described with the normal distribution, safety stock is a function of the standard deviation of the random variable jointly representing two phenomena pertaining to demand variability and supply variability, respectively:

$$ZB = z(p) \cdot \sigma \tag{1}$$

where:

ZB – the safety stock [number of items],

z – the safety coefficient depending on the assumed consumer service level and computed based on the inverse standard normal distribution with the mean 0 and standard deviation 1,

p – the assumed customer service level,

σ – the standard deviation of stock [number of items].

To determine σ, the probability distribution is used of a variable defined as the sum of the random number of values of a random variable with known probability distribution (Krzyżaniak, Cyplik 2007).

Then:

$$\sigma = \sqrt{\sigma_p^2 \cdot T + \sigma_t^2 \cdot P^2} \tag{2}$$

where:
σ – the standard deviation of stock [number of items],
σ_p – the standard deviation of demand reserve per time unit [number of items],
T – the number of time units consumed for supply execution,
σ_t – the standard deviation of the reserve of time units consumed for supply execution,
P – the mean demand per time unit [number of items].

The values of T appearing in formula (2), and thus the values of σ_t are dimensionless, as they define numbers of time intervals (Silver *et al.* 1998). The formula is usually treated as obvious use of statistics, not requiring derivation (see, e.g. Silver *et al.* 1998, Bozarth, Hanfield 2007, Krzyżaniak, Cyplik 2007). However, such an approach blurs the very essence of the phenomenon of safety stock maintenance.

Safety stock may be determined another way, with use of the random variable defined as the difference between aggregate demand (X) over the supply execution period and the actual supply (Y). The standard deviation relates to the difference of two independent variables both of which are expressed in terms of items of goods involved. We assume the random variables to be independent and subordinated to the normal distribution. The variance of the stock may then be expressed as the sum of variances of individual factors:

$$\sigma^2 = \sigma_P^2 + \sigma_D^2 \tag{3}$$

where:
σ_P – the standard deviation of demand [number of items],
σ_D – the standard deviation of supply volume [number of items].

Over the supply execution time T_d, the average number of collections from a warehouse, with intensity P_i, and standard deviation σ_{pi} equals T_d/T_p; thus, the standard deviation of demand over the same period is:

$$\sigma_P = \sigma_{pi} \cdot T_p \cdot \sqrt{\frac{T_d}{T_p}} = \sigma_{pi} \cdot \sqrt{T_p \cdot T_d} \tag{4}$$

where:
σ_{pi} – the standard deviation of demand intensity [items per day],
T_d – the supply execution time [days],
T_p – the gap between two consecutive collections [days].

With respect to supplies it may be assumed that the change in the value of a supply is proportional to the change in time between two consecutive supplies, with supply intensity equal to demand intensity.

The standard deviation of a change in stock caused by a change in supply time is:

$$\sigma_D = \sigma_T \cdot P_i \tag{5}$$

where:

σ_T – the standard deviation of supply time [days],

P_i – the demand intensity, that is demand per unit of time [items per day].

Substitution of (4) and (5) into (3) yields:

$$\sigma = \sqrt{\sigma_{pi}^2 \cdot T_d \cdot T_p + \sigma_T^2 \cdot P_i^2} \tag{6}$$

Upon the substitutions $T = T_d/T_p$ and $T_p = 1$, we get $\sigma_{pi} = \sigma_p$ and $\sigma_T = \sigma_t$, and the formula takes form (2). In view of the above derivation, also the mathematical model used to describe the of warehouse stock becomes comprehensible. Inn such approach to computing security stock, the relations among random variables are considered only. The statistical phenomenon corresponding to the formula just derived consists in the random value of aggregate demand being compared with artificially changed quantity of supplies. If the sum of random demands exceeds the random value of supplies, it is assumed that a deficit has occurred at the warehouse and the demand is not satisfied. Everything is correct as long as random numbers are concerned, but the actual situation at the warehouse is different.

First, consecutive random values of demand are subtracted from the warehouse stock. Only when the current demand exceeds the current warehouse stock, is the demand not satisfied. It may then be assumed, as in the statistical model, that the given supply has failed to satisfy not only a given individual demand, but also the aggregated demand. It may happen, though, that the next demand is satisfied, because it happens to be lower than the warehouse stock balance. Second, only after the supply reaches the warehouse, may satisfying the next demand start. Third, supplies accumulate at the warehouse, so that uncollected quantities increase the probability of satisfying the next demand. None of these phenomena is included in the statistical model adopted. This in turn means that it is necessary to check whether the theoretical formula reflects the warehouse reality.

In the model adopted, the service level may be determined from formula (1). There is:

$$ZB = z(p) \cdot \sigma = F_{0,1}^{-1}(p) = F_{m,\sigma}^{-1}(p) - m, \tag{7}$$

where:

$F_{0,1}$ – the distribution function of the standard normal distribution with the mean 0 and standard deviation 1,

$F_{m,\sigma}$ – the distribution function of the normal distribution with the mean m and standard deviation σ,

$m = P \cdot T$ – the average supply quantity.

Hence, the service level is expressed as:

$$p = F_{m,\sigma}(m + ZB) = P(x < m + ZB). \tag{8}$$

The quantity p is the probability of the aggregate of collection sequence being less than the supply plus the safety stock. According to formula (8), in the absence of safety stock, the service level so understood equals $F_{m,\sigma}(m) = 0.5$. This means that, on average, a half of aggregate collections could not be fully satisfied. Thus the service level $p = 0.5$ would be achievable in practice only if the entire demand for a given supply were satisfied on a one-off basis. This does not correspond to the conditions of stock increases and decreases at the warehouse. Before the stock delivered in a single supply is exhausted, multiple collections adding up to the variable X will be performed. This means that in the sequence of collections, perhaps few collections at the end of the queue would not be executed because of too low stock.

However, this approach is only one of the two interpretations of the service level encountered in the literature (e.g., Krzyżaniak, Cyplik 2007), which may lead to confusion. According to (8), under the first approach, with the quantity of supply D, the service level is expressed by the following formula:

$$p_1(D) = F_{m,\sigma}(D) \tag{9}$$

Using this indicator to assess the service level may be seen as the manifestation of concern for a supply to be large enough to satisfy the aggregate demand of all customers to which the supply has been allotted. This significantly increases the probability that any individual customer has received the demanded goods, provided that increased customer service cost is accepted.

Under the other approach to service level, the probability is determined that the demand has been satisfied, that is the probability that an attempt at goods collection is successful. The service level is computed as the ratio of the demand satisfied and total demand over a given period (Krzyżaniak, Cyplik 2007). The unsatisfied demand may be computed as the integral over the interval from the supply quantity D to plus infinity of the function defined as 1 less the distribution function of the normal distribution with the mean m and standard deviation σ:

$$\begin{aligned} p_2(D) &= 1 - \frac{\int\limits_D^\infty (1 - F_{m,\sigma}(x))\, dx}{\int\limits_0^\infty (1 - F_{m,\sigma}(x))\, dx} \\ &= 1 - \frac{\int\limits_D^\infty (1 - F_{m,\sigma}(x))\, dx}{m} = \frac{\int\limits_0^D (1 - F_{m,\sigma}(x))\, dx}{m} \end{aligned} \tag{10}$$

The two presented ways to understand the service level are so different that the same value may be obtained for a positive or negative safety stock, depending on the approach selected (Tibben-Lembke 2009). A negative stock just means that the supply quantities should be reduced in relation to the mean value, so that the probability of servicing an individual customer is appropriately low. Certainly, if supplies were not reduced (i.e., if there were 0 security stock), then the probability of servicing an individual customer would be higher than assumed.

It might appear that the safety stock corresponding to the service level p_1 is treated as an unavoidable sacrifice on the altar of science, because in practice the safety stock is usually determined in a much less scientific manner, for instance, as 150% of the expected demand (Bozarth, Hanfield 2007). However, instead of using the service level p_2, it is suggested that formula (1) be supplemented with the list of coefficients (relating to supply time, supply cycle duration or forecast time) which are expected to reduce the computed safety level and harmonise it with the conditions at an enterprise (URL). It is suggested that values of those coefficients be identified empirically. This way, the scientific approach would be maintained, while the coefficients would contribute to reducing the burden of the "theoretical" safety stock. However, it is suggested that, instead of relying on safety stock, good communication be maintained with the supplier in order to handle supply variability in practice.

3. Order level

At an actual warehouse, a high service level for an individual customer is achievable even with the zero safety stock. This is possible owing to the fact that non executed collections are kept at the warehouse and – without an intention of the operator – a sort of safety stock is created, anyway. However, warehouse management cannot be based on a single determination of parameters of supply and demand streams, because it is difficult to guarantee any fixed demand for a given product and matching supplies at fixed intervals and quantities. Therefore, a rational solution is to track demand and control supplies accordingly on an ongoing basis in response to changes in demand.

For this reason, an ordering model is used based on what is known as the information level (Bozarth, Hanfield 2007). Supplies are performed on request, that is a supply is ordered when a certain collection reduces the stock below some predefined level known as the order level. Owing to this, temporary changes in demand are compensated by earlier or postponed supplies. The time for sending an order depends on the time T of the response to order, with the time T changeable randomly. It is also in this case that, given demand fluctuations, the average order level is increased by the safety level guaranteeing the predefined service level. Thus the order level is determined according to the following formula (cf. e.g., Bozarth, Hanfield 2007):

$$Z_z = Z_{sr} + ZB = P \cdot T + ZB, \tag{11}$$

where:

Z_z – the order level [number of items],

Z_{sr} – the average order level [number of items].

4. Simulation model

Supply and collection streams under mutual interaction were modelled in the model used to derive formula (6). The graph of the model build of two events is presented in Figure 1.

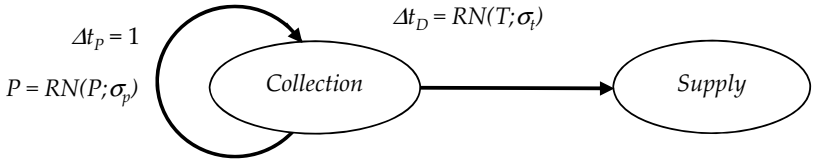

Fig. 1. *Graph of events in the simulation model*

The event *Collection* occurs at unit (fixed) time intervals. It consists in subtracting a random demand quantity P from the stock. If following the subtraction of the current collection the warehouse stock goes below the order level expressed by formula (11), a supply is planned; the supply reaches the warehouse after a random time T. The event *Supply* consists in adding the fixed supply quantity to the stock. The order level equals the initial stock plus supply quantity.

If the current collection is less than the stock, a deficit is registered and the customer collects what is left of the stock. This serves the basis for estimating three service levels in the simulation model. The first level, corresponding to ratio (9), expresses the probability of satisfying the entire demand before the next supply:

$$r_1 = 1 - \frac{L_b}{L_p} \tag{12}$$

where:

r_1 – the service level estimated according to formula (9),

L_b – the number of occurrences of demand non satisfied completely before the next supply,

L_p – the number of supplies.

The second level, corresponding to ratio (10), expresses the probability of satisfying the assumed demand:

$$r_2 = 1 - \frac{Q_b}{Q_p} \tag{13}$$

where:

r_2 – the service level estimated according to formula (10),

Q_b – the difference (expressed as the aggregate number of items of goods) between the demand and stock balance,

Q_p – the aggregate notified demand.

Finally, the third level expresses the probability of satisfying the demand notified by an individual customer:

$$r_3 = 1 - \frac{N_b}{N_p} \tag{14}$$

where:

r_3 – the service level estimated for an individual customer,

N_b – the number of not executed demand orders,

N_p – the number of customer reports for goods.

Figure 2 presents the simulation algorithm.

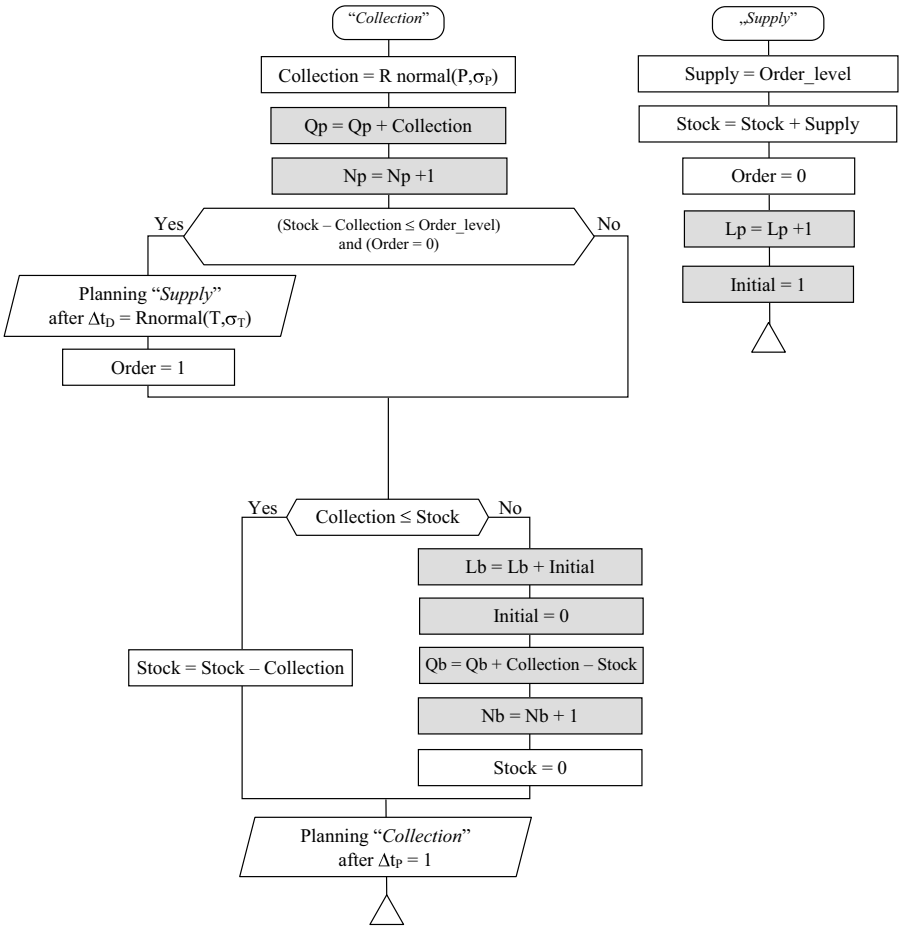

Fig. 2. *Logic flowchart of the simulation model*

5. Example

For the purposes of the example, we assume that a daily demand changes according to the normal distribution with the mean $P = 16$ and standard deviation $\sigma_p = 1$. The time between an order and its physical delivery to a warehouse is governed by the normal distribution with the mean $T = 9$ and standard deviation $\sigma_t = 1.5$. The average order level is increased by safety stock. Supplies were modelled with the value $D = 16 \cdot 9 + ZB$ for various values of ZB.

To examine individual demand, the same model was used, but with a changed demand stream. Collections governed by the normal distribution with the mean $P' = P/16 = 1$ and standard deviation $\sigma_p^{'} = \sigma_p/\sqrt{16} = 0.25$ occurred at fixed time intervals of $1/16$ of a day. They added up to the daily demand equal to that in the basic model.

The model was executed with use of the simulation program, which supports modelling of any process described by a graph and appropriately defined instructions executed at individual nodes of the graph (called cycles) (Tibben-Lembke 2009). For the experiments run, the simulation time $T_{sym} = 10,000$ and number of repetitions $r = 30$ were assumed.

Table 1 quotes control instructions of the simulator for the example analysed (with changed instructions for the additional model given in boxes).

Table 1. *Program controlling the operation of simulator – order level model*

```
<CYKL INI
<informowanie
pobranie.Lb=0
pobranie.Lp=0
pobranie.pierwszy=1
pobranie.Qb=0
pobranie.Qp=0
pobranie.Nb=0
pobranie.Np=0
zamowienie=0
zapas_bezp=0  {od -144 do 80
poziom_zam=144+zapas_bezp
zapas=poziom_zam
*----------------- eksperyment
Tsym=10000
Powtorzenia=30
Czas=0
Pobranie=0
<zlecanie
!pobranie,0.001
#koniec,Tsym
<cykl
*---------------------------
<CYKL koniec
<informowanie
wynik=pobranie.Lb/pobranie.Lp
STOP.r1=STOP.r1+wynik
wynik=pobranie.Qb/pobranie.Qp
STOP.r2=STOP.r2+wynik
wynik=pobranie.Nb/pobranie.Np
STOP.r3=STOP.r3+wynik
Eksp=Eksp+1
<decydowanie
d1:Eksp<INI.Powtorzenia
d2:
<zlecanie
d1:
#INI
d2:
#STOP
<cykl
*---------------------------
<CYKL STOP
<informowanie
r1=1-STOP.r1/INI.Powtorzenia
r2=1-STOP.r2/INI.Powtorzenia
r3=1-STOP.r3/INI.Powtorzenia
<zlecanie
<STOP
<cykl
```

```
*============================
<cykl pobranie
<informowanie
Pobranie=RAND(Pobranie,'norm',16,1)
     Pobranie=RAND(Pobranie,'norm',1,0.25)
Qp=Qp+Pobranie
Np=Np+1
<decydowanie
zam:(INI.zapas-Pobranie<=INI.poziom_zam)
   *(INI.zamowienie=0)
tak:Pobranie<=INI.zapas
nie:Pobranie>INI.zapas
<zlecanie
#pobranie,1
     #pobranie,0.0625
tak:
INI.zapas=INI.zapas-Pobranie
nie:
Lb=Lb+pierwszy
pierwszy=0
Qb=Qb+Pobranie-INI.zapas
Nb=Nb+1
INI.zapas=0
zam:
INI.zamowienie=1
Czas=RAND(Czas,'norm',9,1.5)
!dostawa,Czas
<cykl
*============================
<cykl dostawa
<informowanie
Dostawa=144+INI.zapas_bezp
INI.zapas=INI.zapas+Dostawa
INI.zamowienie=0
Lp=Lp+1
pobranie.pierwszy=1
<cykl
```

Figure 3. presents, for various values of safety stock, results of computations executed according to formulae (9) and (10) and simulation-based estimation of a service level according to formulae (12), (13) and (14), for demand allocated to a time unit and value unit.

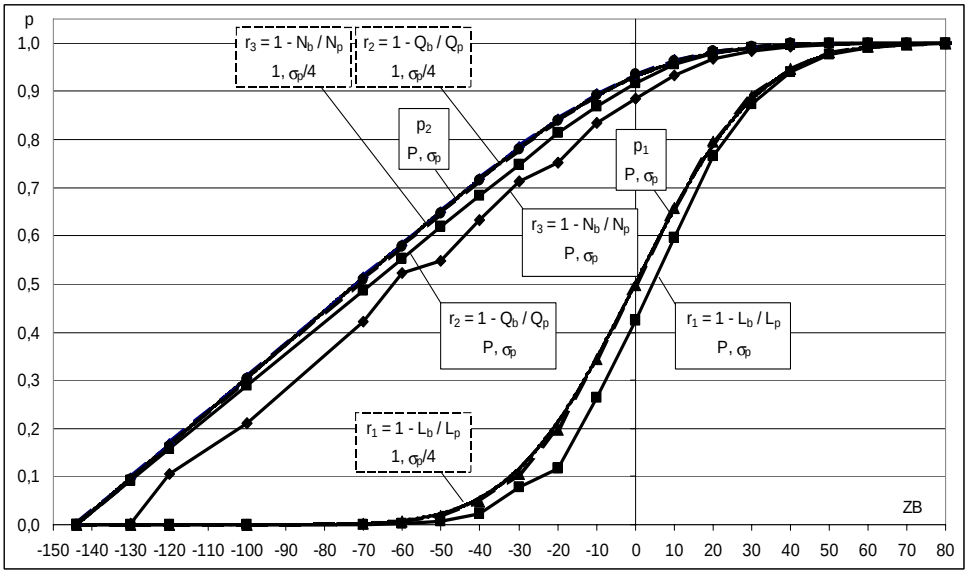

Fig. 3. *Various metrics of service level as a function of safety stock*

Over the entire range of changes in ZB, service levels computed with use of the ratios p_1 and p_2, respectively, are definitely different. The service level expressed with formula (12) is close to theoretical value (9), while that expressed with formula (13) is close to theoretical value (10).

This confirms the conclusion that it is unreasonable to use formula (9) to determine the safety stock, as the stock determined this way is excessive, when compared with the needs of an individual customer. It all the more so, because while considering an individual customer's demand, both metrics of the service level, according to formulae (13) and (14), yield almost equal values and both are almost equal to theoretical value (9). This means that formula (9) describes the service level for an individual customer even if the demand is expressed with respect to a time unit.

If supply quantity equals order level, the service levels determined by simulation are less than theoretical ones. The reason is that in the simulation model – unlike in the theoretical one – delayed supplies are not available and do not contribute to demand satisfaction. The importance of this factor decreases, when the frequency with which customers appear to collect goods supplied increases. If supplies had

fixed quantity larger than order level, the service level would increase, because the time required to satisfy the demand would then increase.

The results of simulation lead to the conclusion that formula (10) describes the service level for an individual customer better and thus is better suited to estimating safety stock created as protection against random fluctuations of supply and demand. Using this formula enables the safety stock to be reduced with the required service level maintained. This is important, as safety stock generates the fixed cost of stock maintaining. Therefore, the cost of stock maintaining may be reduced by not only increasing the cycle (Bonelli 2009), but also realistically estimating safety stock.

Moreover, the use of the simulation model supports the estimation of necessary safety stock for various probability distributions, and not just for the normal distribution, which is commonly used, because the formula for safety stock is relatively simple then.

6. Conclusions

The discussion leads to the conclusion that in the theory of safety stock, the importance of temporary deviations from the mean is exaggerated by requiring that the aggregate demands be satisfied with predefined probability over the order execution period. Indeed, if variables are governed by a normal distribution, then, without safety stock, that is true for a half of instances only.

The high value of safety stock results from the fact that customer service level is assessed with use of the probability of successful servicing all customers ascribed to a given supply. This is convenient for a warehouse operator, as it frees him from tracking fluctuations of supply and demand streams on an ongoing basis, in order to match those streams to each other. It also renders maintaining effective relations with suppliers insignificant. However, it is the customers who pay for this idyll, as they have to accept increased stock service cost.

However, an individual customer only cares for whether his demand is satisfied with a large probability, but he does not care for the satisfaction of the entire customer group. From this point of view, making the supply system ready to satisfy the needs of the entire customer group only generates unnecessary cost, which the customer has to incur. Thus, the formula for service level as it appears in the literature does not reflect the customer's interests and requires excessive stock.

It has also been proved here that the random characteristics of supply and demand streams have a minor effect only on the creation of safety stock. The safety stock quantity is rather a function of supply service skills required to re-establish or maintain broken or suspended relations with suppliers.

Safety stock is necessary for maintaining uninterrupted satisfaction of demand in the period necessary for supply service to restore relations with suppliers broken due to a failure which could not been factored in the statistical analysis of supply process. Therefore, safety stock should be determined not by statistical analysis, but so as to reflect the skills of supply service and way in which relations with suppliers are established.

References

A close call. Logistics Europe, February 2004, pp. 20.

Bonelli M. 2009. *Cut your inventory by 20%*. CSCMP Supply Chain Comments, Vol. 43.

Bozarth C., Hanfield R.B. 2007. *Wprowadzenie do zarządzania operacjami i łańcuchem dostaw* (Introduction to operations and supply chain management). Helion (Polish-language edition).

Feller W. 1957. *An introduction to probability theory and its applications*. John Willey & Sons Inc., New York.

Krzyżaniak S., Cyplik P. 2007. *Zapasy i magazynowanie, tom I* (Stocks and Warehousing, Vol. I). ILiM, Poznań (in Polish).

Okulewicz J. 2000. *Modelowanie symulacyjne współdziałania podsystemów logistycznych* (Simulation modelling for cooperation of logistics subsystems). Prace Wydziału Zarządzania Politechniki Częstochowskiej, Częstochowa, pp. 144–152 (in Polish).

Tibben-Lembke R.S. 2009. *Setting safety stock using a fill rate*. URL `http://www.business.unr.edu/faculty/rtl/463/FillRate.pdf` (accessed April 27th 2009).

Silver E.A., Pyke D.F., Peterson R. 1998. *Inventory Management and Production Planning and Scheduling*. Wiley, New York.

URL `www.inventoryops.com`.