The effect of demand distributions on the performance of inventory policies

SONJA KUHNT & WIEBKE SIEBEN
FAKULTÄT STATISTIK
TECHNISCHE UNIVERSITÄT DORTMUND
44221 DORTMUND

Inventory policies provide instructions on how to calculate stock levels and order quantities. This paper examines the effect of the demand distribution on the performance of several well-known policies. Their performance is compared in terms of achieved service levels. As a conclusion from these comparisons, a nonparametric demand model is proposed to be used in inventory policies.

Keywords: Inventory policies, nonparametric demand model, applications in logistics

1. Introduction

Inventory control policies offer support in establishing the amount to be stocked such that future demand will be met best. Financial aspects as well as the necessity of providing good service are relevant when choosing the optimal stock level and replenishment quantity. For the general concepts compare [Silv. 1998] or [Temp. 2006] and the literature mentioned therein.

We restrict our investigations to the demand of one item at a time in a single-item single-level periodic review setting with order-up-to policies. This means that the stock level is updated at the end of each period and a decision is made, whether the inventory should be refilled up to the order-up-to level S using an inventory policy (see Fig. 1).
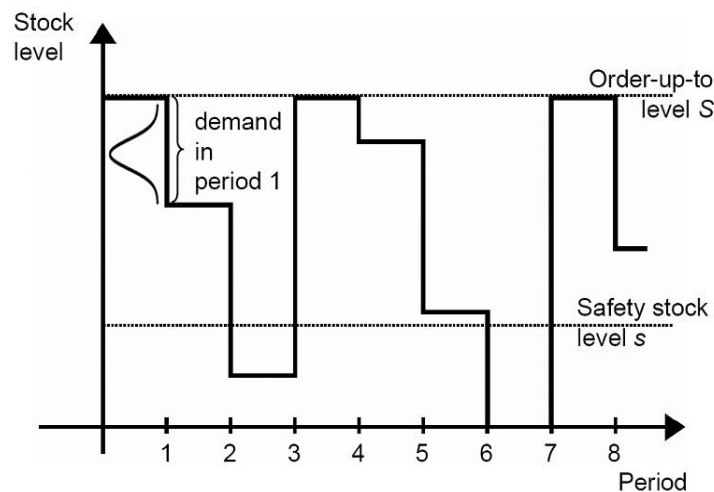

Figure 1: Setup

Along with a vast body of literature, there exist many different inventory policies, some of them being very specific and elaborate, taking into account lots of variables and conditions. Most inventory policies make at least some assumptions on the demand, for example the common assumption of i.i.d. normal demand, but these assumptions are usually not checked in practice.

The aim of this study is to demonstrate the effect of valid and violated demand model assumptions on the performance of inventory policies. To study pure demand model effects, the lead time is set zero, which means that the inventory is refilled instantaneously. This is of course not a realistic assumption as replenishing takes some time and is usually random but it serves our purpose.

First stochastic approaches with respect to the optimal stock level and the modelling of inventory situations were made in the Fifties of the last century. Arrow *et al.* [Arro. 1951] developed optimal (*s,S*)-policies for a static and a dynamic inventory model. Both models make assumptions concerning the demand distribution and take shortage costs into account. In the dynamic model the development of the inventory level is regarded as a Markov chain. Shortly later, in 1953 Dvoretzky *et al.* [Dvor. 1953] modelled the inventory situation as a two-person zero-sum game with loss functions and decision rules. Regarding the optimality of inventory policies Scarf [Scar. 1960] pointed out, that under linear holding and shortage costs an optimal policy is always of (*s,S*)-type. Based on these findings Iglehart [Igle. 1963] developed an approximatively optimal policy for the infinite horizon problem.

Today, many publications focus on the optimality of ordering policies in inventory models. These models are often quite elaborate and specific. Cetinkaya and Parlar [Ceti. 2004] propose the optimal stationary base-stock policy for a periodic review inventory model with independent identically distributed demand, non-linear shortage costs and finite planning horizon. Gallego and Hu [Gall. 2004] present optimal policies for finite and infinite planning horizons in a periodic review inventory model with finite capacity, in which the demand and supply processes both are regarded as Markov chains. Hollier, Mak und Lai [Holl. 2002] develop an optimal (*s,S*)-policy for a continuous review inventory model with compound Poisson distributed demand. If the demand exceeds some critical value and if the inventory level is below some critical value, the demand will be transformed into a replenishment order directly while all other demands will be met from stock.

Some other approaches expand the deterministic "economic order quantity" (EOQ) model [Harr. 1915]. Yu [Yu 1995] develops a robust EOQ model containing the demand, ordering costs and holding costs as random variables. Presman and Sethi [Pres. 2004] introduce a policy for a continuous review inventory model with a demand distribution with a deterministic and a stochastic component. This policy reduces to the EOQ formula, if the stochastic component is missing.

Other publications take the difference between demand forecasts and actual demands into account. Heath and Jackson [Heat. 1994] present their "Martingale Model of Forecast Evolution". Toktay and Wein [Tokt. 2001] transfer this idea to an inventory model with finite capacity. Abels [Abel. 1999] suggests to supplementary order the upper confidence boundary for the median of the forecast errors in addition to the forecasted demand.

There is obviously a wide variety of inventory models and policies. In practice and particularly in large inventories with many items, model assumptions on the demand distribution of every single item cannot be checked due to costs and efforts this would cause. It is therefore of outstanding importance to examine practical inventory policies with few model assumptions with respect to their goodness of service (achieved service level) and compare achieved service levels if even these few assumptions are violated. Overall an inventory policy is required that copes well with many potentially underlying demand distributions. This paper compares inventory models and ordering policies with respect to their performance in simulated and real inventory situations. Also, the impact of highly frequent zero-demand periods, which are a common problem in inventory practice, on the resulting service level is investigated.

The paper starts with a discussion of service levels as measures of good service in Section 2. Afterwards, in Section 3, six inventory models and their ordering policies are presented. In Section 4 simulated inventory situations and real data with a challengingly high proportion of zero-demand periods are described. In Section 5, inventory policies are applied and differences between target and achieved service levels compared. Section 6 provides a summary and discussion of the main results.

2.   Service levels

Comparing inventory policies by their service performances in different inventory situations requires the use of a service measure. The three commonly used service measures are the α, β and γ service level.

In general, service levels measure the ability to deliver demanded goods. They reflect the relationship between demand, met demand and unmet demand and so they are powerful tools when comparing inventory policies. The measures are estimates for unknown theoretical probabilities of stockouts related to the assumed underlying inventory model. The achieved α *service level* is the empirical proportion of periods with fully met demand compared to all periods. The achieved β *service level* is the empirical proportion of the demand met in the period of its occurrence to the total demand. The β service level provides information on the amount missing, while the achieved α service level just notes whether it came to shortages or not. For the γ *service level*, as a third measure of service, different definitions exist. Generally it measures not only the size of the amount missing but also the time of being unable to deliver. In cases of no backorders but lost sales and a lead time assumed to be zero, as we will have later, the γ service level is the same as the β service level. For further details on these service measures see [Schn. 1981].

As every period without demand is a period with fully met demand, an article with 80 percent zero demand periods achieves an α service level of 80 percent, when it is not kept in stock, i.e. the stock level for this article is always zero. If the article is kept in stock, i.e. the stock level is greater than zero, then the achieved α service level will be at least 80 percent. Given a target α service level of 90 percent, the possible difference between the achieved α service level and the target α service level is smaller for an article with many periods without demand than for an article with less periods without demand. Nevertheless, the α service level is an useful tool for judging the ability to deliver demanded articles. An article demanded only once a year maybe needs not to be kept in stock but can be ordered with a certain lead time when requested. The consequences of stockouts need to be weighed up against costs and efforts of having the particular article in stock. If the accommodation of demand is considered compulsory, the α service level can be set so high, that the target α service level cannot be achieved by periods without demand only.

In contrast, the β service level ignores periods without demand. Even so, it is not more suitable to capture the ability to meet the demand than the α service level. If an item is ordered just once a year and due to this not in stock, then none of the demanded amount can be met in the period when the demand occurred. In this example the achieved β service level for a whole year equals zero. It does not reflect, that the inventory level of zero is appropriately chosen most of the time. We therefore concentrate on the α service level as a measure for the ability to meet the demand.

3. Inventory Policies

Inventory models are based on assumptions concerning the nature of the treated inventory situation. Using these models, inventory policies are instructions on how to calculate the optimal stock level and order quantity. Cost or service level considerations can be used to find an optimal policy. These two approaches are often equivalent and both leave space for the users' subjectiveness, as typically not all costs are known and need to be guessed and there is no objective way to determine the appropriate target service level.

We discuss six different (1,S) inventory policies as special cases of (s,S) strategies. Two of the six policies contain fixed safety and base stock levels up to which the inventory is replenished in the beginning of each period. The first inventory policy "P1" is based on the assumption of an i.i.d. normal demand. For the second inventory policy "P2" an i.i.d. gamma distribution is assumed. Inventory policies "P3" and "P4" have fixed safety stock levels but the base stock levels are updated in each period by using past and recent demand observations in a one-step-ahead prediction of an autoregressive integrated moving average (ARIMA) model. The fixed safety stock levels are determined via the variance of the white noise process (P3) or the α quantile of the shortages that occur when no safety stock is kept (P4) respectively. The remaining inventory policies "P5" and "P6" have safety and base stock levels, which are in each period calculated from the last demand observations. For inventory policy P5 again i.i.d. normal demand is assumed, but in contrast to P1 the safety stock and order-up-to level are recalculated every period. Policy P6 uses the empirical distribution function as estimate for the unknown demand distribution, which is updated every period. In all policies the lead time is set zero and no costs but service level constraints are considered. Permitting backorders or not therefore makes no difference.

3.1. Policies with fixed base stock and safety stock levels

Inventory policies P1 and P2 are **parametric policies** that provide rather simple formulae for calculating the optimal order-up-to level S^*. Like in the static AHM-model [Arro. 1951] it is assumed that the **demand** x_t **at any point in time** t **is independent identically distributed** with distribution function F. Here, S^* is not calculated by minimizing costs but by giving a target α service level to be achieved. Thus, S^* is the optimal order-up-to level that satisfies the equation

$$P(x_t \leq S^*) = F(S^*) = \alpha$$

In case of P1, normally distributed demand with expectation μ and variance σ^2 is assumed, and so S^* equals

$$S^* = \Phi^{-1}(\alpha) \cdot \sigma + \mu$$

where Φ is the standard normal distribution function. This order-up-to level can be split into base stock level μ and safety stock level $\Phi^{-1}(\alpha) \cdot \sigma$. This is the usual textbook formula.

In case of P2 the distribution function F is that of a gamma distribution and S^* equals the α quantile of F. The order-up-to level S^* can be split into a base stock level given by the median $\widetilde{x}_{0.5}$ and a safety stock level $S^* - \widetilde{x}_{0.5}$.

3.2.   Policies with fixed safety stock levels and updated base stock levels

The assumption of independent demand observations might, however, often be too far from reality. In such a case, **correlated demand**, for example, can be modelled by an ARIMA model (P3 and P4). These policies, too, are **parametric**.

If the underlying process of correlated demand is an ARIMA process, the demand x_t in period t can be described by

$$\left(1 - \sum_{j=1}^{p} \varphi_j B^j\right)(1-B)^d (x_t - \mu) = \left(1 - \sum_{j=1}^{p} \theta_j B^j\right) \cdot \varepsilon_t$$

where $\theta_1, ..., \theta_q, \varphi_1, ..., \varphi_p$ are constant, B the backshift operator, μ the mean, and ε_t a white noise process with variance σ_ε^2. Under certain assumptions the order-up-to level

$$S_t^* = \hat{x}_t + \Phi^{-1}(\alpha) \cdot \sigma_\varepsilon$$

is optimal [John. 1975], with $\hat{x}_t = E(x_t \mid x_{t-1}, ..., x_1, \hat{\varepsilon}_{t-1}, ..., \hat{\varepsilon}_1)$. Hence for inventory policy P3, the optimal order-up-to level S_t^* in period t is given by the sum of the one step ahead prediction $\hat{x}_t$ and the α quantile of the distribution of ε_t. Here S_t^* can be split into the base stock level $\hat{x}_t$ and the safety stock level $\Phi^{-1}(\alpha) \cdot \sigma_\varepsilon$.

Inventory policy P4 has it's base stock level calculated as policy P3 but it's safety stock level is given by the empirical α quantile of the distribution of shortages [Abel. 1999], that occur when only the one-step-ahead prediction is supplied.

3.3.   Policies with updated base and safety stock levels

Policies with updated base and safety stock levels use all available information for the decision on the amount to be ordered: in every period the estimates are recalculated using the latest demand observations.

For the following policies it is assumed that the demand is independent and identically distributed with a distribution function F, like for P1 and P2. The optimal order-up-to level S^* for a given target service level α is then again given by

$$P(x_t \leq S^*) = F(S^*) = \alpha \Leftrightarrow S^* = F^{-1}(\alpha).$$

Policy P5 is an updated version of policy P1, based on the assumption of normal demand. The base stock level μ_t in period t is estimated by $\overline{x}_t$, the mean demand during periods 1 to t-1. And the safety stock level is estimated by $\Phi^{-1}(\alpha) \cdot \hat{\sigma}_t$, which is the product of the α quantile of a standard normal distribution and the updated estimated standard deviation which together form the optimal order-up-to level S_t^* in period t.

As an alternative to the previously described parametric inventory policies we introduce a **non-parametric policy** (P6) resting on the assumption of **independent identically distributed demand** observations. The unknown demand distribution is estimated by the empirical distribution function. Its order-up-to level S_t^* in

period t can be split into base stock level $\widetilde{x}_{0.5,t}$, the updated median, and safety stock level $\widetilde{x}_{\alpha,t} - \widetilde{x}_{0.5,t}$, the difference between S_t^* and the median.

With these two policies the inventory will be replenished at the beginning of every period up to the optimal order-up-to level S_t^*, unless the inventory level at the end of the previous period is greater than S_t^*.

It is also possible to take only a fixed number of recent demand observations to estimate F_t^{-1}, such that the α quantile of the empirical demand is calculated in a moving window.

4. Inventory situations

The introduced inventory policies are based on assumptions with respect to the demand distribution. The effect of violated distributional assumptions in real inventory situations can not be determined, as the true distribution of the demand is of course unknown. Applying the inventory policies to real and simulated inventory situations and then comparing the achieved service levels, however, can reveal how the policies react to violated model assumptions. We base our analysis on the following simulated inventory situations.

First, demand series are simulated from normal and gamma distributions, as they are widely assumed in inventory practice. Parameters for a simulation of 500 series with length 52 from the gamma distribution were generated by estimation from real inventory situations. Generated this way, the simulated gamma distributed demand series will resemble real demand series, if the assumption of gamma distributed demand is true. Normal distributions are symmetric around the expectation and negative realizations always occur with a certain probability. As demand can not be negative, parameters for simulating normal distributed demand are chosen in a way that negative demand has only a small probability to be generated. So, 500 series from normal distributions are simulated with expectation 1000 and random variance between 1 and 250.

In addition to the simulated inventory situations real inventory situations are considered. The data comes from a coal-mining company. It consists of inventory information on 11830 items containing, amongst others, the pronouncement date of demand, the desired delivery date and the date of the actual delivery. The items are primarily spare parts, as for instance screws and tubes, paint and varnish, instruments, machines and parts of them, but also work clothes, detergents, diesel fuel, sanitary equipment, fire extinguishers, First-Aid devices, personal hygiene devices, coffee, tea and biscuits. The available data is recorded for one year with respect to the date of the actual delivery although the date of the desired delivery comes closest to the demand, the variable of interest. So all deliveries of one year are at hand but not all desired deliveries.

From the data it cannot be seen whether the delivered amount corresponds to the amount demanded or whether it has been split up into part-deliveries. However, potential part-deliveries are merged by summing up all delivered amounts of one article with the same desired delivery date and so the daily demand for one article is formed.

There is a large number of days without demand, especially most Saturdays and Sundays. We set the period length to one week, which reduces the number of zero demand periods compared to a period length of one day. This choice appears to be reasonable as demand is rarely met in the week of pronouncement and delays of up to two years occur. However, despite this choice of period length the demand is still intermittent (see table 1).

We group the items with respect to their achieved α service levels when the stock level is always kept zero, which is just the same as the zero demand frequency for each item. Applying ordering policies to these groups may reveal the influence of zero demand frequencies on achieved service levels. Items with their first quartile of

zero demand frequency	number of items
up to 10%	616
10% - 20%	263
20% - 30%	298
30% - 40%	345
40% - 50%	376
50% - 60%	472
60% - 70%	723
70% - 80%	1056
80% - 90%	1701
90% - 95%	2060
at least 95%	3920

Table 1: Zero demand frequency in weekly demand data

demand greater than zero have less than 25 percent periods without demand. 1001 items fall into this category, they form group 1. Group 2 consists of all items with their first quartile equal to zero but their second quartile greater than zero. It contains 897 items. Analogical the 1785 items are assigned to group 3. They have their second quartile equal to zero and their third quartile greater than zero. The remaining 8139 items build group 4. All items in this group have zero demand in at least 75 percent of all periods.

None of the inventory models presented in Chapter 3 can be regarded as solely underlying in the real inventory demand series. χ^2-goodness-of-fit tests lead to the rejection of the hypothesis of normal, exponential, poisson and lognormal distributed demand for all or nearly all items. Only for 588 items from group 1 the hypothesis of gamma distributed demand could not be rejected. But for most of these 588 items ARIMA-processes could neither be eliminated as underlying processes. The plausibility of underlying ARIMA-processes has due to the vast amount of data only been checked by tests for correlated residuals. Most of the test results do not indicate inadequacy of the fitted ARIMA models.

5. Performance of the inventory policies

Most of the described inventory policies contain unknown parameters which need to be estimated before the policies can be applied to the inventory situations. To estimate the unknown parameters in P1 and P2, we cut the demand series into two halves. The required parameters are for each series estimated from the first half, and then the policies are applied to the second half of the demand series.

When applying inventory policies P3 and P4 to data, ARIMA models need to be fit. The order of these models, the parameters $\theta_1,...,\theta_q,\varphi_1,...,\varphi_p$ and the variance σ_ε are unknown. For policy P3, order and variance are estimated from the first half of the demand series. The parameter estimates are updated in every period of the second half when applying the policies to the data and the one step ahead prediction for the next period is calculated. If the resulting order-up-to level is less than the stock level at the end of the previous period, nothing is replenished. For policy P4, model order and variance are estimated from the first quarter of the series. Then, in every period of the second quarter of the series, the parameter estimates are updated and the one-step-ahead prediction is supplied in the following period. The empirical α quantiles of the shortages in the second quarters of the series form the safety stock levels. Inventory policy P4 is then applied to the remaining second halves of the demand series.

For policies P5 and P6 the demand series are halved to provide for comparability although no parameters need to be estimated from the first halves. Like all other policies they are then applied to the second halves of the series.

The inventory policies are applied to the simulated demand series and the demand series from groups 1 to 4 (of size 500 each) with target service levels of 50, 70 and 90 percent. Inventory policies P3 and P4 differ in the calculation of the safety stock level but are the same in the calculation of the base stock level. Comparing the service levels achieved with P3 and P4 lets us exclude P4 from further examinations, because its performance is even worse than P3. The differences between the achieved and target service levels are displayed by boxplots (see Fig.2, Fig.3, Fig.4). Ideally the median of such a performance boxplot should be zero and the range should be small. When analyzing these figures, one needs to keep in mind, that the policies are applied to only half of the demand series and that not every second half of the demand series of an item in group 4 for example has a zero demand proportion of at least 75 percent as the whole series has.
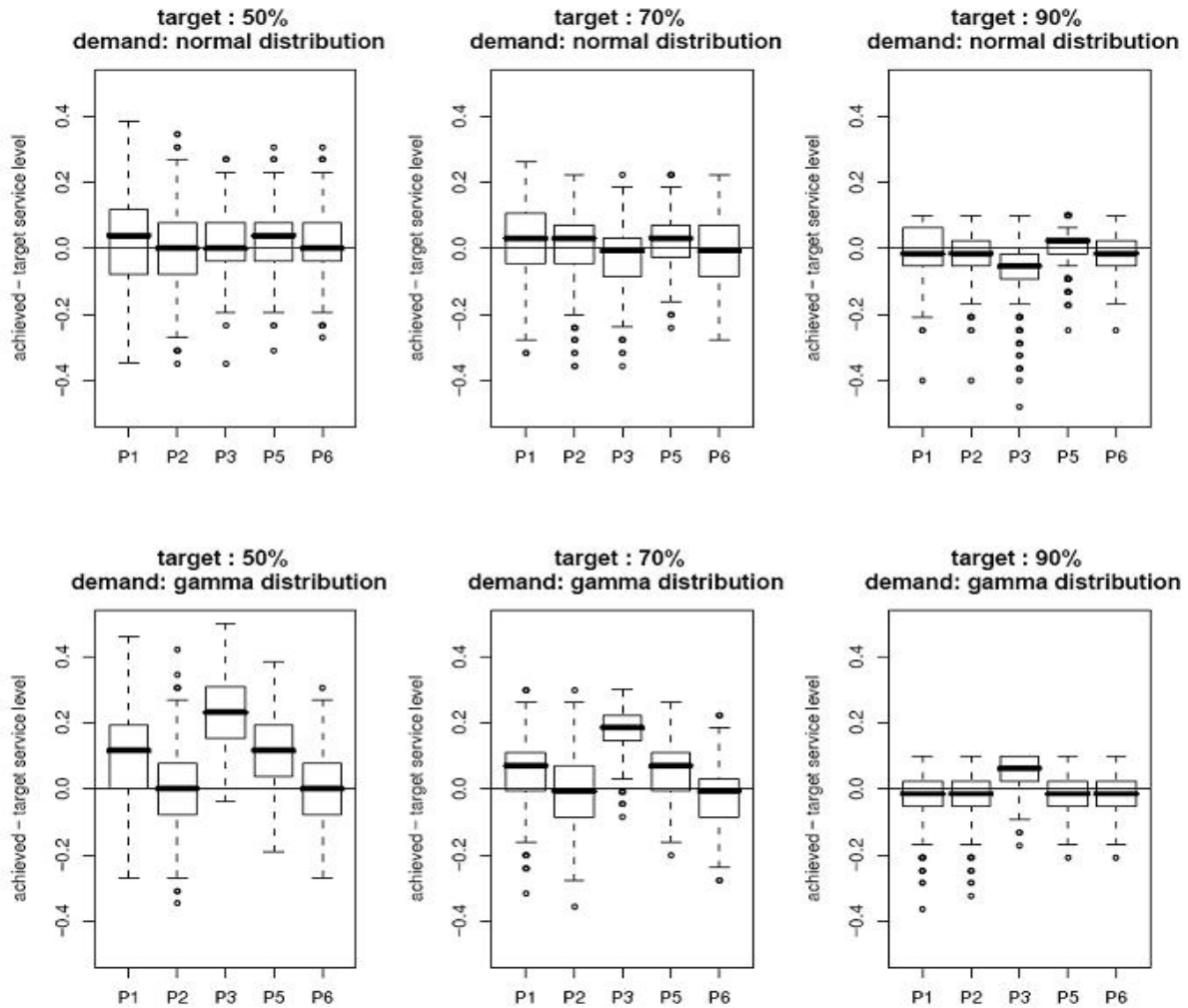
**Figure 2: Differences between target and achieved service level,
gamma and normal demand**

The comparison of target and achieved *α* service and safety stock levels under the different inventory policies leads to several conclusions.

Firstly, we consider the *impact of model or distribution assumptions on achieved service levels*. Neither valid nor violated demand model assumptions imply better or worse performance. Policies with different distribution or model assumptions lead to different achieved service levels. But surprisingly, violated assumptions do not necessarily lead to higher absolute differences from the achieved to the target service levels.

This can be seen, when comparing the performance of P1 and P2 on the simulated normal data. Although policy P2 has a gamma demand assumption which is violated it performs better than policy P1 which has the valid normal demand assumption. On the simulated gamma data policy P2 again performs better than policy P1. In this case the policy with valid demand model assumption leads to better performance than P1 with violated demand model assumption.
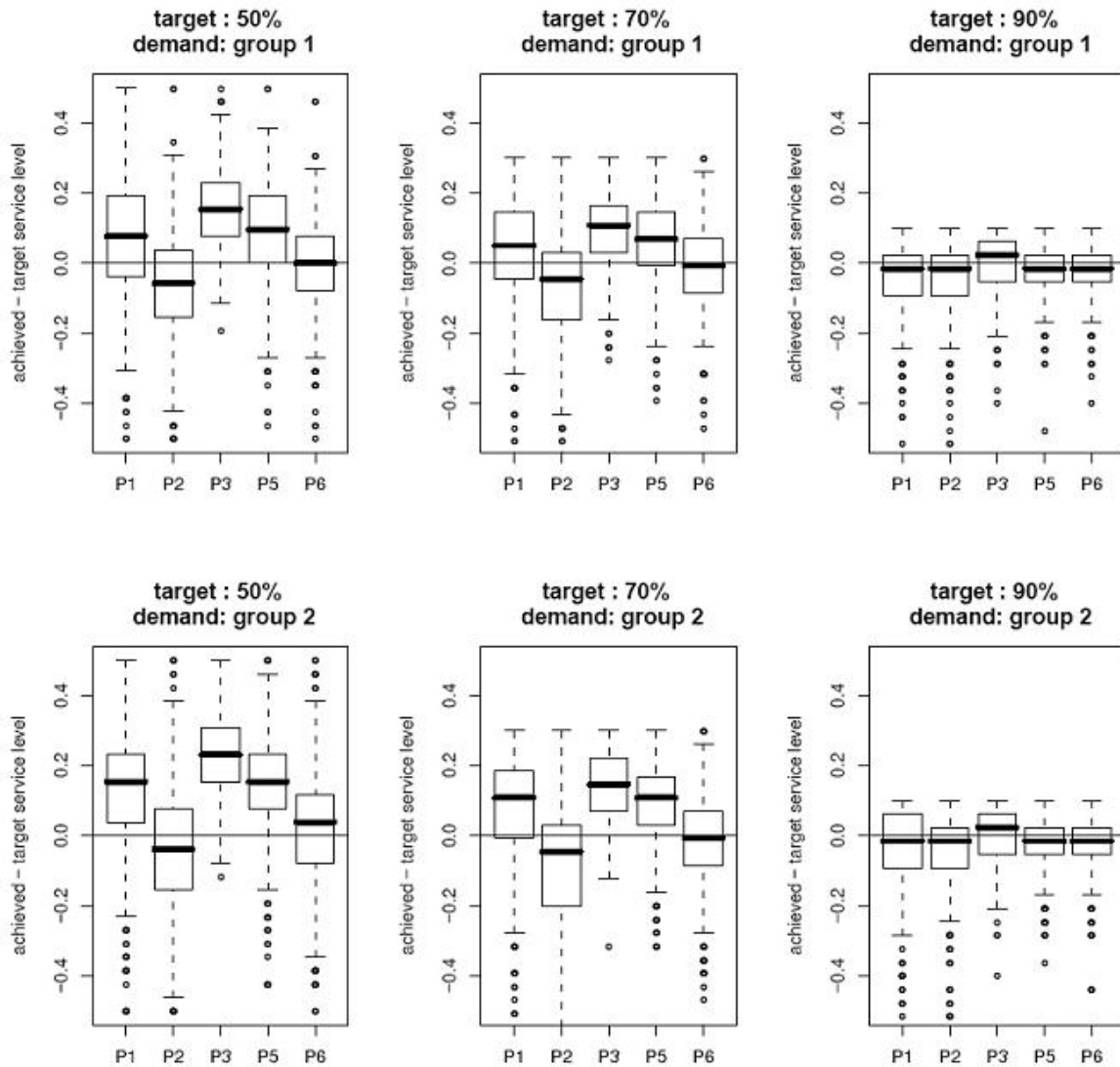
**Figure 3: Differences between target and achieved service level,
group 1 and group 2**

Secondly, the performance gets better with rising target service level, the *impact of the target service level*. Along with rising α service levels, the differences between inventory policies with respect to meeting the given service level become more and more negligible. And so does the choice of an inventory policy.

And thirdly, there is an *impact of zero demand proportions on achieved service levels*. If the proportion of zero demand periods exceeds the target α service level, the achieved service level will exceed the target service level. Besides this, inventory policy P1 (assuming normal distribution) leads to rising achieved α service levels with rising zero demand proportions even when the zero demand proportions do not exceed the target service levels. So if the past rate of zero demand of an item is higher than the desired service level or close to it, the target service level is already achieved by zero demand periods and it might be unnecessary to store it in the future. These findings suggest to first determine the past rate of zero demand and then decide on the necessity of an inventory policy.
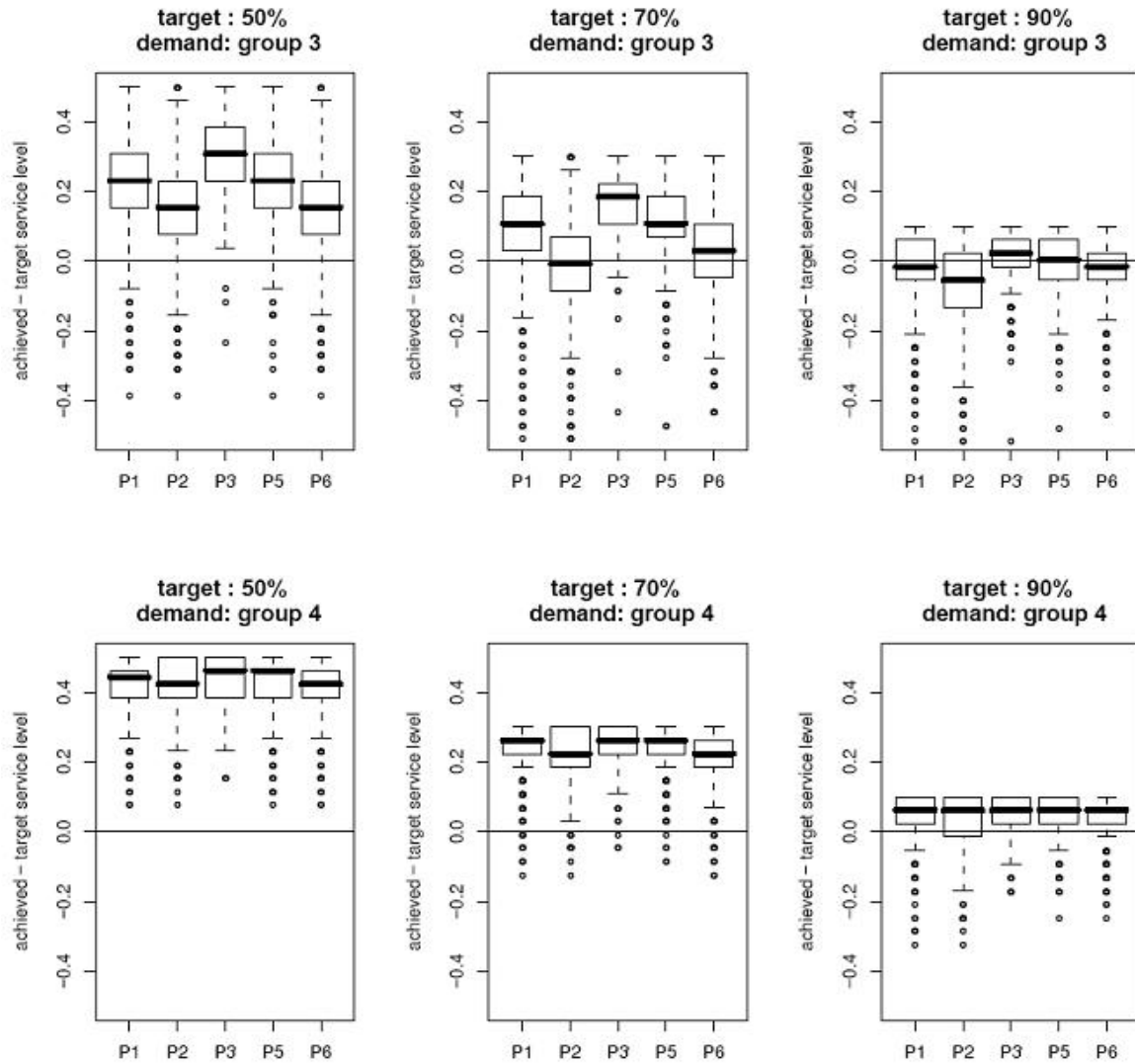
**Figure 4: Differences between target and achieved service level,
group 3 and group 4**

For groups 1 to 4 of the real inventory situation the policies listed in table 2 turned out to meet the target α service levels best. If just one policy is to be chosen for all items in an inventory, then the nonparametric policy P6 can be recommended.

demand	target α service level		
	50%	70%	90%
group 1	P6	P6	P6
group 2	P6	P6	P6
group 3	P2 oder P6	P2 oder P6	P3 oder P6
group 4	P1 oder P6	P1 oder P6	P2 oder P6

Table 2: Policies with best achieved α service levels

6.   Summary

In practical applications inventory policies are needed for automated decisions on optimal inventory levels and order quantities. The majority of inventories contains large numbers of items, making extensive individual decision processes impossible.

Different inventory policies have been proposed in the literature, varying with the related inventory model and model assumptions. In most inventories, however, the validity of model assumptions cannot be checked individually for every item. Therefore inventory policies are investigated in this study with respect to their performance in simulated situations with valid and violated demand model assumptions. Additionally the adequacy of the policies is examined in a real inventory situations. The investigated real inventory data show a high proportion of zero demand periods. The inventory models compared in this study cover policies with

- i.i.d. normal and gamma demand distributions
- autoregressive integrated moving average modelled demand.

The different model and distributional assumptions led to changes in the achieved service levels, but violated assumptions did not necessarily lead to bigger differences between achieved and target service levels. It turned out, that achieved service levels were influenced by zero demand proportions. The achieved service level cannot be smaller than the zero demand proportion and so the achieved service level of rarely demanded items can exceed the target service level without being kept in stock. As a result of this study, none of the demand models could be regarded as appropriate for all considered real inventory situations. The updated nonparametric policy performs best on the whole. Further, neither valid nor violated demand model assumptions lead to better or worse performance. Consequently, we propose that updated distribution free demand estimates should be used for inventory management.

Bibliography

[Abel. 1999]   Abels, H.: DISKOVER II, Ganzheitliche Bestimmung von Sicherheitsbeständen. Berlin Heidelberg New York: Springer.

[Arro. 1951]   Arrow, K.J., Harris, T., Marschak, J.: Optimal inventory policy. Econometrica **19**, 250-271.

[Ceti. 2004]   Cetinkaya, S., Parlar, M.: Computing a stationary base-stock policy for a finite horizon stochastic inventory problem with non-linear shortage costs. Stochastic Analysis and Applications **22**, 589-625.

[Dvor. 1953]   Dvoretzky, A., Kiefer, J., Wolfowitz, J.: The inventory problem: II. case of unknown distributions of demand. Econometrica **21**, 450-466.

[Gall. 2004]   Gallego, G., Hu, H.C.: Optimal policies for production/inventory systems with finite capacity and Markov-modulated demand and supply processes. Annals of Operations Research **126**, 21-41.

[Harr. 1915]   Harris, F. W.: Operations Cost. Factory Management Series, Chicago: Shaw.

[Heat. 1994]   Heath, D.C., Jackson, P.L.: Modelling the evolution of demand forecasts with application to safety stock analysis in production/distribution systems. IIE Transactions **6**, 17-30.

[Holl. 2002]   Hollier, R.H., Mak, K.L., Lai, K.K.: Computing optimal (s,S) policies for inventory systems with a cut-off transaction size and option of joint replenishment. International Journal of Production Research **40**, 3375-3389.

[Igle. 1963]   Iglehart, D.L.: Optimality of (s,S) policies in the infinite horizon dynamic inventory problem. Management Science **9**, 259-267.

[John. 1975]   Johnson, G.D., Thompson, H.E.: Optimality of myopic inventory policies for certain dependent demand processes. Management Science **21**, 1303-1307.

[Pres. 2004]   Presman, E., Sethi, S.P.: Stochastic inventory models with continuous and poisson demands and discounted and average costs. Working Paper, The University of Texas at Dallas, Richardson, TX.

[Scar. 1960]   Scarf, H.: Optimality of (s,S) policies in the dynamic inventory problem. Mathematical Methods in the Social Sciences. Stanford, California: Stanford University Press, K.J. Arrow, K. Karlin, P. Suppes, eds.

[Schn. 1981]   Schneider, H.: Effect of service-levels on order-points or order-levels in inventory models. International Journal of Production Research **19**, 615-631.

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

[Temp. 2006]   Tempelmeier, H.: Inventory Management in Supply Networks. Norderstedt: Books on Demand.

[Tokt. 2001]   Toktay, L.B., Wein, L.M.: Analysis of a forecasting-production-inventory system with stationary demand. Management Science **47**, 1268-1281.

[Yu 1995]   Yu, G.: Robust economic order quantity models. European Journal of Operations Research **100**, 482-493.