SERVICE LEVEL BASED CAPACITY RATIONING PROCEDURE FOR MAKE-TO-ORDER MANUFACTURING SYSTEMS

M. Sharifyazdi and M. Modarres*

Department of Industrial Engineering, Sharif University of Technology
Tehran, Iran
sharifyazdi@mehr.sharif.edu - modarres@sharif.edu

*Corresponding Author

(**Received: July 26, 2006 - Accepted in Revised Form: January 18, 2007)**

Abstract We extend a heuristic method within the framework of "dynamic capacity apportionment procedure" (DCAP) to allocate an existing capacity among the classes with different profit contributions. In general, DCAP is applied when some capacity shortage exists and can not be enhanced in short - run. Our proposed approach is constructed for a make - to - order manufacturing system that produces a variety of products while experiences a burst of demand in excess of capacity. Although, a higher level of profit can be gained by accepting more orders from higher priority classes at the expense of rejecting some or all of orders of lower priority classes, it may result in elimination of an existing market segment. To avoid this case, which practitioners are very much concerned about it, we propose an approach by maintaining a desired minimum service level for each product class. This method of rationing policy maximizes the expected profit by discriminating product classes while meeting the individual product service level targets set by the management. We also highlight the managerial implications of such a result and identify possible avenues for further research.

Keywords Capacity Management, Demand Management, Revenue Management, Customer/ Product Service Level, Make-to-Order Manufacturing

چکیده دراین مقاله برای تخصیص ظرفیت یک سیستم تولیدی بین گروههای مختلف مشتریان، روشی ابتکاری در چارچوب مبانی مدیریت درآمد ارائه می شود. کاربرد این روش در مورد سیستمهای تولید سفارشی است که باکمبود ظرفیت مواجه می شوند. این سیستمها می توانند تولیدات متعددی داشته باشند. هنگامی که تقاضا بیش از ظرفیت باشد، برای مقابله با کمبود، پذیرش سفارش مشتریان کالاهای پرسودتر و عدم پذیرش سفارش‌های کم‌سودتر، به درآمد بیشتر منتهی می شود. لیکن این امر منجر به از دست دادن قسمتی از بازار نیز خواهد شد. لذا در رویکرد ارائه شده، با منظور نمودن یک سطح حداقل برای پذیرش سفارش‌های هرگروه از مشتریان – که به آن سطح خدمت مطلوب گفته می‌شود – از این امر جلوگیری می‌گردد. در روش پیشنهادی برای تخصیص ظرفیت، ضمن قائل شدن تفاوت بین گروههای مختلف مشتریان و تامین هدف سطح خدمت هر گروه، درآمد کل حداکثر می‌شود. کاربردهای مدیریتی این روش نیز بررسی و برای تشریح این رویکرد مثالی ارائه می‌گردد. این مدل، در چارچوب فرایند پویای تخصیص ظرفیت قرار دارد. هدف مدل تخصیص ظرفیت محدود موجود به گروههایی از مشتریان است که سود حاصل از انجام سفارش‌های آنها با هم متفاوت است.

1. INTRODUCTION

In both manufacturing and service sectors, effective management of supply and demand is vital for maximizing profit and at the same time expanding the market and making the customers more satisfied,. Mass customization, electronic commerce, virtual supply chains and other managerial developments have forced firms to reduce delivery time and improve reliability, simultaneously (Barut and Sridharan [1]). These trends increase the importance of capacity/demand management as a central function to the resolution of the conflict between manufacturing and

marketing (Sridharan [2]).

Managing capacity and allocating it between various competing sources of demand is neither a new subject nor unique for service industries and manufacturing firms. In the service management literature, one can find many works regarding the issue of allocating capacity between competing classes of customers (demand) for airlines, hotels, rental car agencies, (Smith, Leimkuhlcr and Darrow [3]). As Balakrishnan, Sridharan and Patterson [4] discuss, the objective of the capacity allocation problem in make - to - order (MTO) and assemble-to-order (ATO) manufacturing systems are similar to that of service industries. In both cases the main issue is how to allocate some perishable asset, which is production capacity, among different classes of products (or customers) in order to maximize the overall profit. Similarly, MTO manufacturing firms need to establish a capacity management policy in order to solve short - run capacity allocation problems, when demand exceeds the capacity. To ensure operational coordination between marketing and manufacturing, it is vital to exploit the available capacity in the most efficient and effective way. The potential revenue loss caused by unused customizable capacity is the same as unsold airline seats. By acceptance or refusal of orders, a limited capacity is allocated among multiple product classes with different profit contributions. This also can be considered as yield management problem (Harris and Pinder [5]). Clearly, there are rewards (e.g., increased profits) and penalties (e.g., lon - term impact on market share) associated with accepting or rejecting orders for each product, respectively.

Given the total demand is greater than the available capacity, it may be tempting to conclude that all a firm needs to do is to satisfy the demand of the most profitable class and ignore less profitable classes when the objective is to maximize profit. This is what a typical dynamic capacity apportionment procedure (DCAP) (Barut and Sridharan [1]) problem does. However, from a strategic point of view in order to avoid losing an important market segment, it may be important to maintain a certain minimum service level for all products in the product mix. That is our main motivation in this research to develop a method which maintains some prescribed service level for each product in a short - term capacity allocation problem faced by make-to-order manufacturing firms encountering excess of expected total demands in comparison with capacity. Actually, what makes this work distinguished from the previous ones in literature is incorporating this strategic objective in the model. Although this model can not be found in literature, some have recommended it, see Kimes [6], Weatherford, Bodily [7], Barut and Sridharan [1].

By applying as well as modifying the DCAP concept, we develop a heuristic to maximize the expected profit while meeting the individual product service level targets set by the management. We assume the company has a distinct pricing policy and the customers purchasing behavior is not affected by this policy. Our model considers differentiated products, nested booking classes, static pricing strategy, dynamic allocation, and also single-period assignment.

In the next section, we briefly review the literature relevant to the problem. In Section 3, we present the heuristically modified DCAP. To conclude the paper, we discuss the managerial implications of such a result and identify possible avenues for further research in this vitally important area of inquiry.

2. LITERATURE REVIEW

To plan make-to-order manufacturing systems, there exists a variety of methods, concepts and approaches in literature, see Stevenson et al. [8]).There are two major decision levels in make-to-order firms, the job entry level and the job release level. At the job entry level, customer enquiries are processed, and delivery dates and prices are quoted to customers. At the job release level, decisions are made regarding which jobs should be released to the shop floor so that processing can commence (Hendry, Kingsman and Cheung [9]).

The Capacity rationing problem, which lies in the job entry level, has received considerable attention in service operations, (Kimes [6]), especially in the context of yield management (revenue management) for airline and hotel industries. The

basis of revenue management is an order acceptance/refusal process that integrates marketing, financial, and operations functions to maximize revenue from pre-existing capacity (Harris and Pinder [5]). Revenue management originates from service (especially airline) industry. At the heart of airline revenue management lays the seat inventory control problem. This problem concerns the allocation of the finite seat inventory to the demand that occurs over time before the flight is scheduled to depart. In order to decide whether or not to accept a booking request, the opportunity costs of losing the seats taken up by the booking have to be evaluated and compared to the revenue generated by accepting the booking request. Moreover, a booking request that creates the highest possible revenue for the airline should never be rejected whenever a seat is available, not even when the number of seats appointed to this type of passenger by the booking control policy has been reached. In fact, any passenger should be allowed to tap into the capacity reserved for any other lower valued type of passenger. This is the concept of nesting and should be incorporated into the booking control policy. In our research, we use these two concepts to develop our model for capacity coordination in MTO firms.

While previously considered primarily as a tool of service operations, revenue management has considerable potential for manufacturing operations. MTO manufacturing firms share the environmental characteristics of companies in which yield management practices has been successfully employed, such as fixed capacity, perishable resource and uncertain demand (Barut and Sridharan [1]). However, prior works on applying the revenue management concept for short - run capacity management in MTO manufacturing environment is limited. Sridharan [2] provides a comprehensive contrast between the capacity allocation problem in manufacturing and the perishable asset revenue management (PARM) problem well developed in the service operations literature. Citing the example of high-fashion apparel industry, Balakrishnan, Sridharan and Patterson [4] propose a single-period rationing model when demand is stochastic. Focusing on the short - term capacity allocation problem faced by a class of make - to - order manufacturing firms encountering expected total demands in excess of capacity, they use a decision - tree analysis to develop a simple policy that may be used to dynamically allocate capacity for two classes of products. Sridharan and Balakrishnan [10] correct a weakness identified in the previous model (Sridharan and Patterson [4]) about rationing policy under certain demand and capacity situations, and extendes the single-period model to the multi - period case with demand uncertainty. They also use a decision tree based approach. Although this model is still limited to two classes of products, the multi - period case allows modeling customer orders with due dates and earliness and tardiness penalties. The scope of these studies has been limited to the two - product class case. These models do not yield optimal solution when more than two fare classes are considered. In contrast our model presents a multiple - product capacity rationing model for managing capacity in MTO manufacturing systems when demand exceeds capacity. The most important similarity among these models and ours is using decision tree based approach, which is based on the idea of equating the marginal revenues in the various fare classes. The second is using the concept of nesting. In all these models an order for capacity utilization is always allowed to tap into the capacity reserved for any lower valued type of orders. Barut and Sridharan [1] extend the model presented by Balakrishnan, Sridharan and Patterson [4] and develop a single-period multiple-product class capacity rationing model. Deploying a decision theory based approach; the authors et al. develop a heuristic for short-term constrained capacity allocation to multiple-product classes in make-to-order manufacturing, attempting to maximize profit by discriminating between product classes. This model has been called Dynamic Capacity Rationing Procedure (DCAP). All rationing policies presented above are “dynamic” in the sense that the capacity rationed for the higher profit classes is continually updated during the planning horizon. Patterson, Balakrishnan and Sridharan [11] show that such a policy consistently outperforms an alternate rationing policy that fixes the rationed quantity at the start of the planning horizon and do not update it as time progresses. For a complete research overview on static and dynamic models in revenue management literature

refer to Pak and Piersma [12]. Capacity rationing in all published models, especially current version of DCAP, focusing on maximizing profit by sacrificing the service level (i.e. fill rate) for the lower priority classes. In these models, one might end up rejecting all orders for lower priority product classes, especially when the capacity is very tight. This may result in loosing customers from a particular market segment. With a few exceptions, not much can be found in the capacity management literature, which deals with customers/products service levels. Fransoo and Sridharan [13], however, define demand management as one that is concerned with setting aggregate sales levels and individual product target sales levels so that the available capacity is effectively utilized. Fransoo, Sridharan and Bertrand [14] present a two - tiered hierarchical approach for scheduling production in multi-item single machine systems, facing very high levels of stochastic demand. The nonlinear programming model embeds a profit maximizing capacity coordination heuristic for determining system parameters (i.e., target cycle times and inventory levels) in long - term, subject to capacity and service level constraints. Hopp and Sturgies [15] use queuing theory to develop a method for allocating capacity and quoting manufacturing due dates to achieve a target service level (percent of orders filled on - time). The last three mentioned papers also aim to achieve and maintain a minimum level of service (in terms of fill rates) for each product or customer, although they do not apply the concept of revenue management to address the issue of service levels in capacity coordination problems. However, all these findings are encouraging in the sense that similar decision models can be developed for handling the customers/products service levels in make - to - order manufacturing's order acceptance problems.

3. SERVICE LEVEL BASED CAPACITY APPORTIONMENT PROCEDURE

We develop our model based on the assumptions and environment similar to the ones by Barut and Sridharan [1]. Consider an MTO manufacturing system with one source of fixed capacity capable of producing different types of products. The fixed planning is divided into N periods. Capacity and order size are expressed in terms of total processing time available to produce products. The products portfolio is grouped into L mutually exclusive classes, based on marginal profit contribution per unit of capacity consumed. The customer orders arrive stochastically and each one consists of only one product class and its size is expressed in terms of capacity units needed to fulfill the order. The product class, order size, and due date characterize orders. Orders have independent and identically distributed random variables size.

We assume the product classes are sorted in a descending order according to their profitability. In other words, class 1 is the most expensive class while class L is the cheapest one. Upon receipt of an order, it is either fully accepted or rejected, based on the remaining available capacity, the marginal profit and the predefined preferred service level of product class ordered. We define service level as the percentage of orders (demand) accepted in each product class. Such order acceptation / refusal process is used in a group of make-to-order firms, called "Versatile manufacturing companies". Amaro, et al.[16] used this term to describe those manufacturers which are involved in a competitive bidding situation for every individual order which they receive. It is a dynamic procedure in the sense that each time an order for a lower profitable product class is received, the model analyzes an objective function consist of expected profit minus expected penalty arising from loss of service level of higher profitable product classes. Then the model maintains an optimal portion of the available capacity for yet to fulfill upcoming orders of higher profitable product class (es) other than the class for which the incoming order is. This is called "protection level" in revenue management terminology. If an order size does not exceed the remaining capacity minus the protected capacities, then the incoming order is accepted. Thus, the only time dependent decision variable is the protected capacity (protection level) for yet to respond demand of higher profitable product classes. The value of this decision variable(s) has to be determined such that the objective function (total revenue) is maximized.

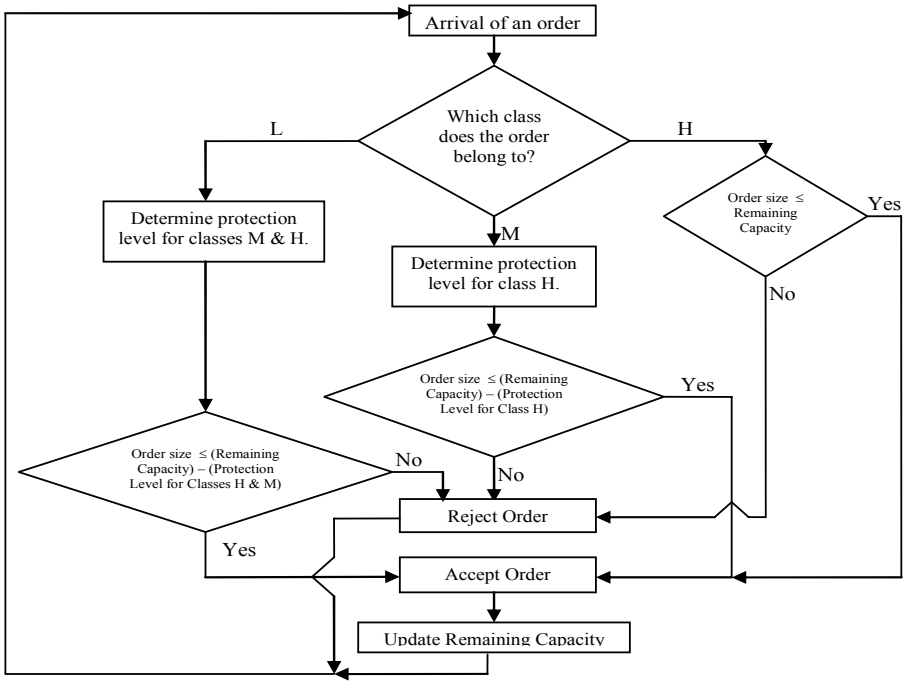

Figure 1. Flowchart representing the decision procedure.

The protection levels are assumed to be nested. It means when c units of capacity is protected for the m^{th} higher profitable product class, they are it is also protected for m-1 higher profitable classes.

To determine the optimal protection level, we compare the value of objective function for discrete non - negative values of protection level in ascending order. The optimal value is the last one, which is greater than the next value to it in objective function.

Orders are assumed to be processed by "first - come first - serve (FCFS) rule. If two orders arrive at the same time, the order of higher profitable class has preference.

The decision process is summarized for a 3 - class case as a flowchart in Figure 1. Classes 1, 2 and 3 are named H, M and L respectively.

We use a decision tree model to calculate the objective value (OV). This model is similar to the one by Barut and Sridharan [1], which is itself an extension of Pfeifer [17] model.

If the incoming order belongs to class j and the protection level for higher profitable classes is set on q_{j-1}^{t}, then four cases could be recognized. These cases are the same as the cases in DCAP (Barut and Sridharan [1]). In each case, the objective value consists of two parts. The first part, which contains expected sales profit, have been previously presented in DCAP, and the second part, which estimates the penalty risen from loss of service level, is the base of originality of this paper.

Case 1.

When

$$\sum_{i=j}^{L} X_{it}^{d\tau} \le q_o^t - q_{j-1}^t$$

and

$$\sum_{i=1}^{j-1} X_{it}^{d\tau} \le q_{j-1}^t$$

In this case, future demand for higher profitable classes (1 to j-1) during time interval [t,dτ] is less than their protection level (q_{j-1}^{t}). Furthermore,

future demand for the class of incoming order as well as for other lower profitable classes (j to L) is less than unprotected capacity (order acceptation limit or $q_o^t - q_{j-1}^t$). According to our notation, probability of occurrence of this case equals $(1-p_t^{d\tau}).(1-\beta_t^{d\tau})$.

Under these conditions, all future orders during [t,dτ] will be accepted and total profit gained is $\sum_{i=1}^{L} P_i X_{it}^{d\tau}$.

There are A_i^t units of demand accepted for class i among total D_i^t units of demand arrived before arrival of current order. So, at the time t, service level for class i is $\frac{A_i^t}{D_i^t}$.If all orders during [t,dτ] ($X_{it}^{d\tau}$) are accepted, then the service level will be $\frac{A_i^t + X_{it}^{d\tau}}{D_i^t + X_{it}^{d\tau}}$ at the time dτ. Hence, the penalty of loss of service level for the whole classes equals $\sum_{i=1}^{L} \pi_i \Pr\left\{S_i > \frac{A_i^t + X_{it}^{d\tau}}{D_i^t + X_{it}^{d\tau}}\right\}$.

Thus, the objective value in this case is as follows.

$$OV = \sum_{i=1}^{L} P_i X_{it}^{d\tau} - \sum_{i=1}^{L} \pi_i \Pr\left\{S_i > \frac{A_i^t + X_{it}^{d\tau}}{D_i^t + X_{it}^{d\tau}}\right\} \quad (1)$$

Case 2.

When

$$\sum_{i=j}^{L} X_{it}^{d\tau} \le q_o^t - q_{j-1}^t$$

and

$$\sum_{i=1}^{j-1} X_{it}^{d\tau} > q_{j-1}^t$$

In this case we study the situation in which future demand of higher profitable classes exceeds protection level but demand of lower classes is less than unprotected capacity. The probability to face this case is $(1-p_t^{d\tau}).(\beta_t^{d\tau})$. Since the demand of higher profitable classes is more than protection level, at least q_{j-1}^t units of capacity will be assigned to the classes 1 to j-1 and all of the remaining orders (belonging to any class) will compete for unprotected part of capacity. Let $O_{[1,j-1]}^a$ and $O_{[j,L]}^a$ be the amount of demand fulfilled for higher and lower profitable classes respectively, from unprotected portion of remaining capacity. As a result of uncertainty, a weighted unit profit is used to estimate total profit gained from both higher and lower profitable classes.

These weighted unit profits are named $\bar{P}_{[1,j-1]}$ and $\bar{P}_{[j,L]}$ for classes 1 to j-1 and j to L respectively and calculated as follows:

$$\bar{P}_{[1,j-1]} = \frac{\sum_{i=1}^{j-1}\left(\frac{P_i}{\theta_{it}^{d\tau}}\right)}{\sum_{i=1}^{j-1}\left(\frac{1}{\theta_{it}^{d\tau}}\right)}$$

and

$$\bar{P}_{[j,L]} = \frac{\sum_{i=j}^{L}\left(\frac{P_i}{\theta_{it}^{d\tau}}\right)}{\sum_{i=j}^{L}\left(\frac{1}{\theta_{it}^{d\tau}}\right)} \quad (2)$$

Hence, total profit could be estimated as $\bar{P}_{[1,j-1]}.[q_{j-1}^t + O_{[1,j-1]}^a] + \bar{P}_{[j,L]}.O_{[j,L]}^a$.

To approximate service level penalty, we assume that service levels are the same for all of the classes 1 to j-1 during [t,dτ] and equals average ratio of fulfilled demand to total demand. A similar assumption is made for classes j to L. That is:

Average service level for classes 1 to j-1 during

$$[t, d\tau] = \frac{\text{fulfilled demand}}{\text{total demand}} = \frac{q^t_{j-1} + O^a_{[1,j-1]}}{\sum_{i=1}^{j-1} X^{d\tau}_{it}}$$

Average service level for classes j to L during

$$[t, d\tau] = \frac{\text{fulfilled demand}}{\text{total demand}} = \frac{O^a_{[j,L]}}{\sum_{i=j}^{L} X^{d\tau}_{it}}$$

So, the overall service level from start of planning time to the time $d\tau$ for each class i could be summarized as:

$$t.\frac{A^t_i}{D^t_i} + (d\tau - t)\frac{q^t_{j-1} + O^a_{[1,j-1]}}{\sum_{i=1}^{j-1} X^{d\tau}_{it}} \quad \text{if } 1 \le i \le j-1$$

and

$$t.\frac{A^t_i}{D^t_i} + (d\tau - t)\frac{O^a_{[j,L]}}{\sum_{i=j}^{L} X^{d\tau}_{it}} \quad \text{if } j \le i \le L$$

In the above relations a weighted average of current service level and estimated future service level as the overall service level is calculated. The weights (t and dτ-t) are the same as lengths of the time intervals that each service level belongs to.

So, total penalty of loss of service level for classes 1 to j-1 could be written as:

$$\sum_{i=1}^{j-1} \pi_i.\Pr\left\{ t.\frac{A^t_i}{D^t_i} + (d\tau - t)\frac{q^t_{j-1} + O^a_{[1,j-1]}}{\sum_{i=1}^{j-1} X^{d\tau}_{it}} < S_i \right\}$$

$$= \sum_{i=1}^{j-1} \pi_i.\Pr \left\{ \sum_{i=1}^{j-1} X^{d\tau}_{it} > \frac{(d\tau - t)\left[q^t_{j-1} + O^a_{[1,j-1]} \right] D^t_i}{S_i.D^t_i - t.A^t_i} \right\}$$

and the same penalty for classes j to L may be estimated as:

$$\sum_{i=j}^{L} \pi_i.\Pr\left\{ t.\frac{A^t_i}{D^t_i} + (d\tau - t)\frac{O^a_{[j,L]}}{\sum_{i=j}^{L} X^{d\tau}_{it}} < S_i \right\}$$

$$= \sum_{i=j}^{L} \pi_i.\Pr\left\{ \sum_{i=j}^{L} X^{d\tau}_{it} > \frac{(d\tau - t).O^a_{[j,L]}.D^t_i}{S_i.D^t_i - t.A^t_i} \right\}$$

Hence, in this case, objective value is summarized as follows:

$$OV = \bar{P}[1,j-1].[q^t_{j-1} + O^a_{[1,j-1]}] + \bar{P}[j,L].O^a_{[j,L]} -$$

$$\sum_{i=1}^{j-1} \pi_i.\Pr\left\{ \sum_{i=1}^{j-1} X^{d\tau}_{it} > \frac{(d\tau - t)\left[q^t_{j-1} + O^a_{[1,j-1]} \right] D^t_i}{S_i.D^t_i - t.A^t_i} \right\}$$

$$- \sum_{i=j}^{L} \pi_i.\Pr\left\{ \sum_{i=j}^{L} X^{d\tau}_{it} > \frac{(d\tau - t).O^a_{[j,L]}.D^t_i}{S_i.D^t_i - t.A^t_i} \right\} \quad (3)$$

Case 3.

When

$$\sum_{i=j}^{L} X^{d\tau}_{it} > q^t_o - q^t_{j-1}$$

and

$$\sum_{i=1}^{j-1} X^{d\tau}_{it} \le q^t_{j-1}$$

Consider the situation in which future demand of lower profitable classes is overtakes unprotected portion of capacity, but for higher profitable classes, demand doesn't reach protection level. Probability of this case is $(p^{d\tau}_t).(1-\beta^{d\tau}_t)$. In this case, all of the orders of higher profitable classes, that occupy $\sum_{i=1}^{j-1} X^{d\tau}_{it}$ units of capacity, will be

accepted, but only a portion of orders in lower profitable classes will be fulfilled. Volume of lower priced fulfilled demand equals the maximum possible value which is the whole unprotected portion of capacity ($q_o^t - q_{j-1}^t$). So, obtained profit during time interval [t,dτ] could be written as $\sum_{i=1}^{j-1} P_i X_{it}^{d\tau} + \bar{P}_{[j,L]}.[q_0^t - q_{j-1}^t]$.

To evaluate achieved service level for higher profitable classes we can use the same method as case 1, because all of the orders in the mentioned classes will be accepted. So, if $1 \le i \le j - 1$, then at the time dτ service level for class i will equal $\frac{A_i^t + X_{it}^{d\tau}}{D_i^t + X_{it}^{d\tau}}$.

However, to estimate gained service level in lower profitable classes, we assume that percentage of demand fulfilled for classes j to L during [t,dτ] is the same and equals the average ratio of fulfilled demand for lower profitable classes during [t,dτ] (as in case 3) which is:

$$\frac{\left(q_0^t - q_{j-1}^t\right)}{\sum_{i=j}^{L} X_{it}^{d\tau}}$$

Hence, for each class i, which lies among classes j to L, final service level at the time dτ could be approximated using the same weighted average method as case 2 as follows:

$$t.\frac{A_i^t}{D_i^t} + (d\tau - t)\frac{\left(q_0^t - q_{j-1}^t\right)}{\sum_{i=j}^{L} X_{it}^{d\tau}}$$

Thus, in the case 3, objective value is summarized as underneath:

$$OV = \sum_{i=1}^{j-1} P_i X_{it}^{d\tau} + \bar{P}_{[j,L]}.[q_0^t - q_{j-1}^t] - \sum_{i=1}^{j-1} \pi_i \Pr\left\{ S_i > \frac{A_i^t + X_{it}^{d\tau}}{D_i^t + X_{it}^{d\tau}} \right\}$$

$$- \sum_{i=j}^{L} \pi_i \Pr\left\{ S_i > t.\frac{A_i^t}{D_i^t} + (d\tau - t)\frac{(q_0^t - q_{j-1}^t)}{\sum_{i=j}^{L} X_{it}^{d\tau}} \right\}$$

$$= \sum_{i=1}^{j-1} P_i X_{it}^{d\tau} + \bar{P}_{[j,L]}.[q_0^t - q_{j-1}^t] - \sum_{i=1}^{j-1} \pi_i \Pr\left\{ X_{it}^{d\tau} > \frac{A_i^t - D_i^t.S_i}{S_i - 1} \right\}$$

$$- \sum_{i=j}^{L} \pi_i \Pr\left\{ \sum_{i=j}^{L} X_{it}^{d\tau} > \frac{(d\tau - t).(q_0^t - q_{j-1}^t).D_i^t}{S_i.D_i^t - tA_i^t} \right\} \quad (4)$$

Case 4.

When

$$\sum_{i=j}^{L} X_{it}^{d\tau} > q_o^t - q_{j-1}^t$$

and

$$\sum_{i=1}^{j-1} X_{it}^{d\tau} > q_{j-1}^t$$

In the last case, available capacity for both higher and lower profitable classes is constrained. So, at least q_{j-1}^t units of capacity will be assigned to higher priced classes and the unprotected portion of capacity will be divided between higher and lower profitable classes. The difference between this case and case 2 is that in case 2 some parts of the unprotected portion of capacity may remain unoccupied but in case 4 all of this unprotected portion will be assigned to orders. Occurrence probability of this case is $(p_t^{d\tau}).(\beta_t^{d\tau})$.

Let w be the fraction of unprotected portion of capacity which is allocated to higher profitable classes. So, (1-w) shows the fraction of unprotected capacity that is filled by lower profitable classes. To approximate w, let d be the fraction of total capacity will be assigned to lower profitable classes, if no protection level exists and

both orders of higher and lower profitable classes arrive on a fixed rate derived from their mean demand.

To determine d, one may simulate this case as a situation, in which two cars are running toward each other with different speeds on a line. One of these cars indicates higher priced classes and is located on the left hand side end of the line and the other one indicates lower profitable classes and lies on the right hand side end of the line. Initial distance between cars is q_0^t (total available capacity). Speeds of the first and the second car are $\sum_{i=1}^{j-1}\left(\frac{1}{\theta_{it}^{d\tau}}\right)$ (mean demand rate for higher profitable classes) and $\sum_{i=j}^{L}\left(\frac{1}{\theta_{it}^{d\tau}}\right)$ (mean demand rate for lower profitable classes) respectively. So, d is the distance of crash point from right end of the line, or in other words, d is the distance that the right hand side car (lower profitable classes) will progress. Let the unit of distance be equal to unprotected portion of capacity ($q_o^t - q_{j-1}^t$). So, d will result, solving the following equation:

$$\left[\sum_{i=1}^{L}\left(\frac{1}{\theta_{it}^{d\tau}}\right)\right].d=\left[\sum_{i=1}^{j-1}\left(\frac{1}{\theta_{it}^{d\tau}}\right)\right].\left[\frac{q_{j-1}^t}{q_o^t-q_{j-1}^t}+1-d\right]$$

and

$$w=\begin{cases}0 & \text{if} \quad d>1\\ 1-d & \text{if} \quad d\le 1\end{cases}$$

Hence

$$w=\text{Max}\left\{0,1-\frac{\left[\sum_{i=1}^{j-1}\left(\frac{1}{\theta_{it}^{d\tau}}\right)\right]\left[\frac{q_{j-1}^t}{q_o^t-q_{j-1}^t}+1\right]}{\sum_{i=1}^{L}\left(\frac{1}{\theta_{it}^{d\tau}}\right)}\right\} \qquad (5)$$

which is different from and more precise than the formula proposed for w in Barut and Sridharan [1]. Thus, it could be considered as an innovation as well.

So, $q_{j-1}^t + w(q_0^t - q_{j-1}^t)$ units of capacity will be assigned to higher profitable classes and $(1-w)(q_0^t - q_{j-1}^t)$. Now, we can summarize total profit of this case as underneath:

$$\bar{P}[1,j-1].[q_{j-1}^t + w(q_0^t - q_{j-1}^t)] + \bar{P}_{[j,L]}.[q_0^t - q_{j-1}^t](1-w)$$

Using a weighted average method similar to cases 2 and 3, service level for a class i is approximated as:

$$t.\frac{A_i^t}{D_i^t}+(d\tau-t)\frac{q_{j-1}^t+w(q_0^t-q_{j-1}^t)}{\sum_{i=1}^{j-1}X_{it}^{d\tau}} \quad \text{if} \quad 1\le i\le j-1$$

and

$$t.\frac{A_i^t}{D_i^t}+(d\tau-t)\frac{(1-w)(q_0^t-q_{j-1}^t)}{\sum_{i=j}^{L}X_{it}^{d\tau}} \quad \text{if} \quad j\le i\le L$$

Thus, the objective value in this case is summed up as:

$$OV=\bar{P}[1,j-1].[q_{j-1}^t+w(q_0^t-q_{j-1}^t)] + \bar{P}_{[j,L]}.[q_0^t-q_{j-1}^t](1-w)$$

$$-\sum_{i=1}^{j-1}\pi_i \Pr\left\{S_i > t.\frac{A_i^t}{D_i^t}+(d\tau-t)\frac{q_{j-1}^t+w(q_0^t-q_{j-1}^t)}{\sum_{i=j}^{j-1}X_{it}^{d\tau}}\right\}$$

$$-\sum_{i=j}^{L}\pi_i \Pr\left\{S_i > t.\frac{A_i^t}{D_i^t}+(d\tau-t)\frac{(1-w)(q_0^t-q_{j-1}^t)}{\sum_{i=j}^{L}X_{it}^{d\tau}}\right\}$$

$$= \bar{P}[1, j-1].[q_{j-1}^t + w(q_0^t - q_{j-1}^t)] + \bar{P}[j, L].[q_0^t - q_{j-1}^t](1-w)$$

$$- \sum_{i=1}^{j-1} \pi_i \Pr\left\{ \sum_{i=j}^{j-1} X_{it}^{d\tau} > (d\tau - t) \frac{q_{j-1}^t + w(q_0^t - q_{j-1}^t) D_i^t}{S_i D_i^t - t A_i^t} \right\}$$

$$- \sum_{i=j}^{L} \pi_i \Pr\left\{ \sum_{i=j}^{L} X_{it}^{d\tau} > (d\tau - t) \frac{(1-w)(q_0^t - q_{j-1}^t) D_i^t}{S_i D_i^t - t A_i^t} \right\} \quad (6)$$

We have shown the four cases mentioned above in a decision tree structure in Figures 1, 2 and 3. Figure 1 illustrates profit in each case and compares profit between the situation in which the protected capacity is q_{j-1}^t and the circumstances in which $q_{j-1}^t + 1$ units of capacity is kept for higher profitable classes. In Figure 2, the same comparison is made for penalty of loss of service level. The total objective function is expected value of OV. Since we subtract objective function for each q_{j-1}^t from that of $q_{j-1}^t + 1$, we will have an incremental objective function ($\Delta OV_t^{d\tau}$). Figure 3 summarizes the incremental objective value which is consist of both profit and service level penalty in all of the cases.

Now, according to the following figures the incremental objective function between time t and dτ could be summarized as follows:

$$\Delta OV_t^{d\tau} = p_t^{d\tau}\left[\beta_t^{d\tau}(1-w)\left[\bar{P}_{[1, j-1]} - \bar{P}_{[j, L]}\right] - (1-\beta_t^{d\tau})\bar{P}_{[j, L]}\right] + p_{it}^{d\tau}(1-\beta_{it}^{d\tau}) \sum_{i=j}^{L} \pi_i \left[\Pr\left\{ \sum_{i=j}^{L} X_{it}^{d\tau} > \frac{(d\tau - t).(q_0^t - q_{j-1}^t - 1).D_i^t}{S_i.D_i^t - t A_i^t} \right\} - \Pr\left\{ \sum_{i=j}^{L} X_{it}^{d\tau} > \frac{(d\tau - t).(q_0^t - q_{j-1}^t).D_i^t}{S_i.D_i^t - t A_i^t} \right\} \right]$$

$$+ p_{it}^{d\tau}\beta_{it}^{d\tau}\left(\sum_{i=1}^{j-1} \pi_i \left[\Pr\left\{ \sum_{i=j}^{j-1} X_{it}^{d\tau} > (d\tau - t) \frac{q_{j-1}^t + 1 + w(q_0^t - q_{j-1}^t - 1) D_i^t}{S_i D_i^t - t A_i^t} \right\} - \Pr\left\{ \sum_{i=j}^{j-1} X_{it}^{d\tau} > (d\tau - t) \frac{q_{j-1}^t + w(q_0^t - q_{j-1}^t) D_i^t}{S_i D_i^t - t A_i^t} \right\} \right] + \sum_{i=j}^{L} \pi_i \left[\Pr\left\{ \sum_{i=j}^{L} X_{it}^{d\tau} > (d\tau - t) \frac{(1-w)(q_0^t - q_{j-1}^t - 1) D_i^t}{S_i D_i^t - t A_i^t} \right\} - \Pr\left\{ \sum_{i=j}^{L} X_{it}^{d\tau} > (d\tau - t) \frac{(1-w)(q_0^t - q_{j-1}^t) D_i^t}{S_i D_i^t - t A_i^t} \right\} \right] \right) \quad (7)$$

As we mentioned before, optimal protection level is the first discrete value, which has a negative incremental objective function.

To calculate the value of incremental objective function for a special q_{j-1}^t, it is necessary to know the probability distribution (density) function of $X_{it}^{d\tau}$. Because, unknown parts of $\Delta OV_t^{d\tau}$, which are $\beta_t^{d\tau}$ (as we know $\beta_t^{d\tau} = \text{pr}\left\{ \sum_{i=1}^{j-1} X_{it}^{d\tau} > q_{j-1}^t \right\}$) and probabilities of $\sum_{i=j}^{L} X_{it}^{d\tau}$ and $\sum_{i=1}^{j-1} X_{it}^{d\tau}$ to be more than a specific value, depend on $X_{it}^{d\tau}$. The value of $X_{it}^{d\tau}$ depends on two random variables, number of orders received during the time interval [t,dτ] (which is shown as Ni) and the size of each order (Yi). Thus, the whole demand of class i during [t,dτ] ($X_{it}^{d\tau}$) may be expressed as a random

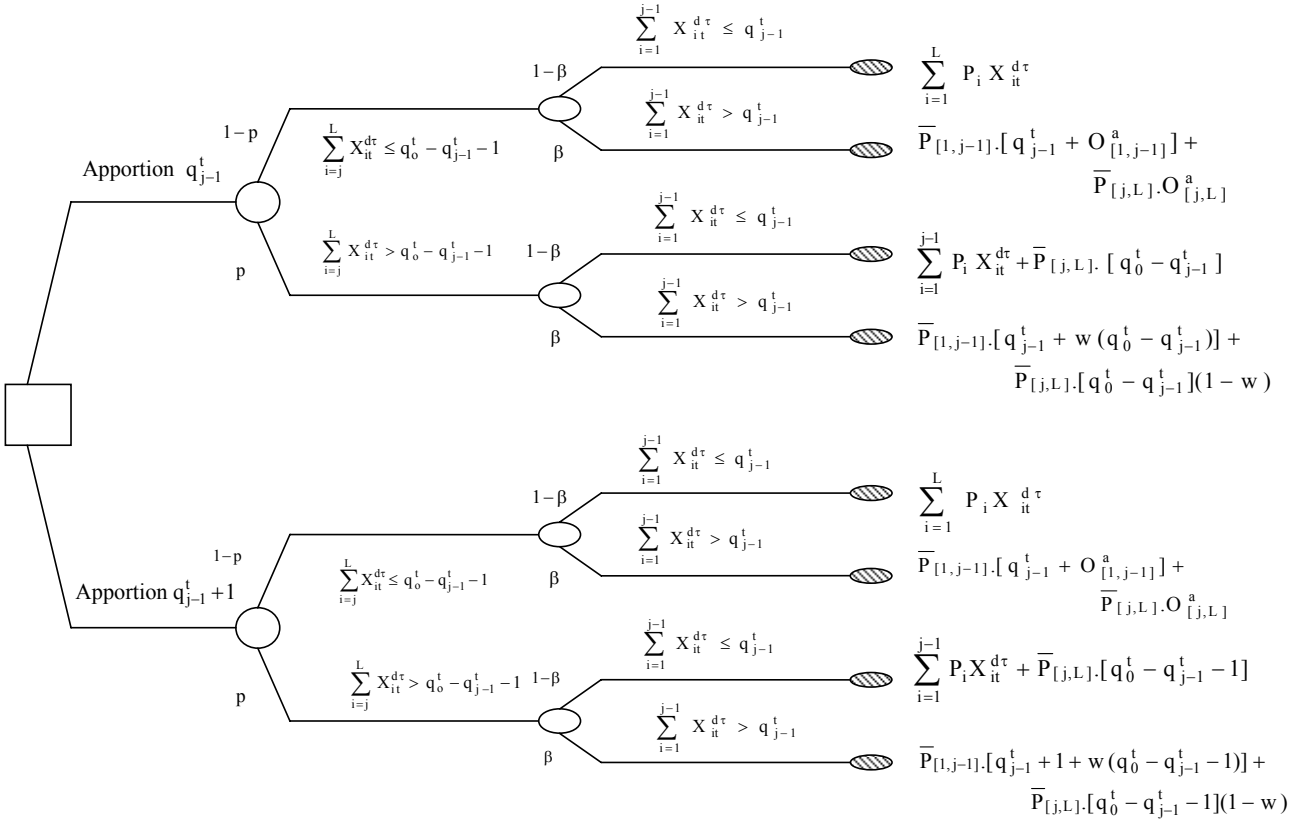

Figure 2. Profit of protecting q^t_{j-1} and $q^t_{j-1}+1$ units of capacity in the four cases.

sum of random variables (Barut and Sridharan [1]):

$$X^{dt}_{it} = \sum_{k=1}^{N_i} Y_{ik}$$

While, the stochastic process of order arrivals is known, the PDF of N_i could be analyzed. For example, when orders arrive in a poisson process, or in other words, when times between order arrivals follow a negative exponential distribution, N_i has a poisson PDF.

So, when probability distribution of Y_{ik}'s is recognized, PDF of $X^{d\tau}_{it}$ could be identified using such methods as conditional distributions, moment generation functions, etc. Das [18] determines PDF of $X^{d\tau}_{it}$ when N_i has a poisson distribution and Y_{ik}'s obey an identical normal distribution. Some other forms of $X^{d\tau}_{it}$ under different conditions are discussed in Hadley and Whitin [19]. Finally, PDF of both $\sum_{i=j}^{L} X^{d\tau}_{it}$ and $\sum_{i=1}^{j-1} X^{d\tau}_{it}$ may be obtained using a convolution on $X^{d\tau}_{it}$'s (Barut and Sridharan [1]).

4. EXAMPLE

In this section to illustrate the proposed approach, we run the procedure for a couple of iterations (order arrivals) to illustrate how it performs.

Consider a case where there are three classes of customers namely H, M and L as described in Figure 1. The planning horizon contains 375

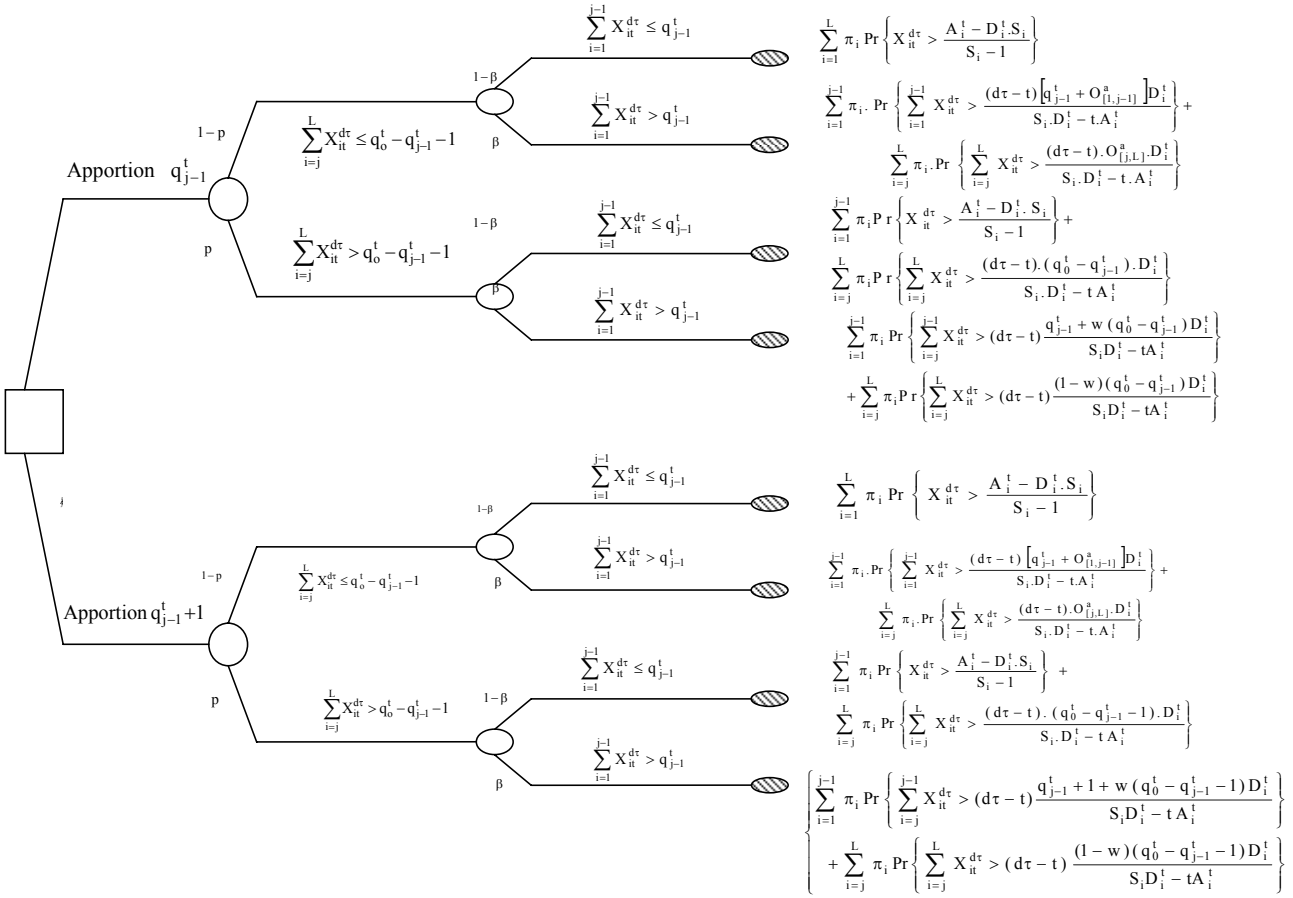

Figure 3. Penalty of loss of service level while protecting q^t_{j-1} and $q^t_{j-1}+1$ units of capacity in the four cases.

periods each period consisting of 8 units of capacity (working hours). So, there are 8 × 375 = 3000 units of capacity available in the whole planning horizon. Orders of each class arrive according to a homogenous Poisson process. Volume of capacity needed to fulfill each order has a Normal distribution, which depends on its class. Parameters of arrival process and order size for each class are listed in Table 1.

Desired service level, penalty of missing it and price rate for each class are listed underneath.

An order from the second class has arrived on the 275th day at the 5th hour. At this time, 275 × 8 + 5 = 2205 hours is passed from beginning of the planning horizon. Assume this order needs 4 units of capacity (4 working hours) to be fulfilled. Let the due date of the arrived order be the 340th period.

Total number of received and accepted orders before this time is shown in the following table as well as total size of received and accepted orders. Achieved service level is also calculated by dividing total size of accepted orders to total size of received orders.

Since the unit of order size (demand) is the same as production capacity, it is shown in terms of working hours. So, the first free working hour available to fulfill demand is 400 + 800 + 1300 = 2500th hour (4th hour on 324th day). As the currently received order has to be done till 340th day, there are only 340 × 8 - 2500 = 220 working hours which can be used to fulfill this order. So, the question is that how many units of capacity (out of 220) should be protected for the first class.

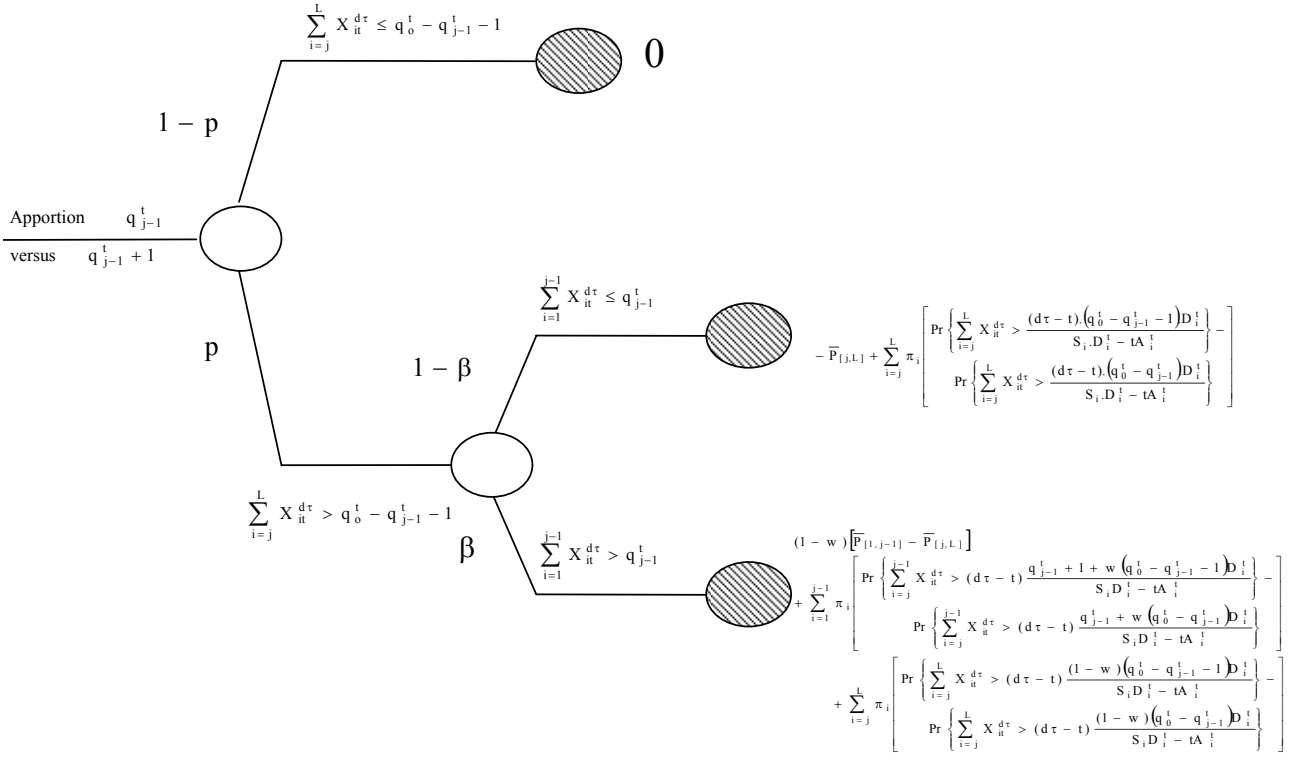

Figure 4. Incremental expected objective value.

If this protection level is not more than 220 - 4 = 116 (4 is size of the order), then this order has to be accepted.

To determine above protection level (q_l^t), a simulation procedure in MATLAB environment is employed. This procedure calculates ΔOV for each integer q_l^t from 0 to 220. The first value of q_l^t which makes ΔOV negative, will be the optimal value of q_l^t. Figure 5 illustrates ΔOV as a function of q_l^t. In this example, ΔOV is positive, until it reaches 106 and becomes - 0.0034. Hence, 106 units of capacity should be kept for the first class.

The unprotected part of capacity is 220 - 106 = 114 and the size of received order is 4. Therefore, the incoming order is accepted. In the second class, the number of received and accepted orders, the total size of received and accepted orders and achieved service level update to 321, 181, 1444, 804 and 0.5568 respectively.

Decision about acceptation/rejection of next orders can be made through a similar process.

5. CONCLUSION

In this paper we developed an approach for capacity rationing problem faced by MTO manufacturing firms expecting total demand in excess of capacity by considering service level. Rejecting an order caused by capacity constraint may have a hidden effect on the arrival rates, due to the negative effect of word - of - mouth that spreads rapidly among customers. In long term, this will result in a considerable market share reduction. By applying, as well as modifying the DCAP concept, our proposed heuristic rationing policy considers both short term and long term profit of the manufacturing firms. It maximizes the expected profit by discriminating between product classes while meeting the individual product

TABLE 1. Order Size and Arrival Rate Parameters.

Class i	Mean order size μ_i	Standard deviation of order size σ_i	Arrival rate λ_i
1: H	3.5	0.9	0.07
2: M	4.5	1.125	0.09
3: L	6	1.5	0.12

TABLE 2. Service Level and Expense Parameters.

Class i	Desired service level S_i	Penalty of losing service level π_i	Price rate P_i
1: H	0.65	200	1
2: M	0.65	230	0.7
3: L	0.65	260	0.5

service level targets set by the management. There are many examples of such capacity rationing problems in MTO situations. Some of them are mentioned in Harris and Pinder [5]. Consider a large repair facility that repairs large industrial transformers. Repairs frequently require both custom work and standard components. In this situation, while some transformers are on routine scheduled maintenance, while others are "emergency" orders due to failure, short-term capacity is fixed and revenues from unused capacity are lost. Other examples are custom sports apparel manufacturers, manufacturers that supply gift stores, paper and plastic dinnerware suppliers, etc.

For further research one can work on evaluating performance of the model to reveal considerable improvement in individual products service level, while total profit has been decreased smoothly. The work done so far is limited to a single period or two products case. This should be extended to address the multiple products case or the multi - period case simultaneously. Thus, increasingly realistic models can be developed for handling complex manufacturing situations. Similarly, it may be valuable to develop advanced dynamic rules for rationing capacity in highly constrained manufacturing systems. Furthermore, modifying the service level based capacity rationing policy in the way that could consider a mix of alternatives in the case of shortage such as subcontracting, substituting products and partial deliveries may be valuable in improving demand management and, hence, worth investigating.

Evaluating incremental objective function for different types of PDF's of both order sizes and inter-arrival times between order arrivals can also be useful to ease computations through execution of the algorithm.

6. NOTATION

The symbols used in this model are summarized as follows:

T	Total time of planning horizon
N	Total number of periods in the planning horizon
τ	Amount of time in each period
t	Arrival time of an order
d	The period in which the order is due to perform
L	Total number of classes
j	The class corresponding to the arrived order
q_0^t	Total available production capacity at time t
q_i^t	Protection level for classes i to 1 (i most profitable classes) at t
$X_{it}^{d\tau}$	Random Variable indicating demand for class i during time interval [t,dτ] (from arrival time of current order to the end of planning horizon)
P_i	Unit profit in class i
$\left(\frac{1}{\theta_{it}^{d\tau}}\right)$	Expected value of demand for class i during time interval [t,dτ]

TABLE 3. Service Level and Expense Parameters.

Class i	Number of Received orders	Number of Accepted orders	Total size of received orders	Total size of accepted orders	Achieved service level
1: H	220	130	770	400	0.5195
2: M	320	180	1440	800	0.5556
3: L	420	230	2520	1300	0.5159

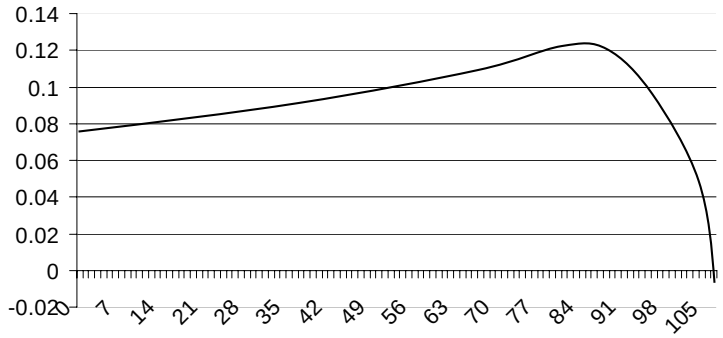

Figure 5. Curve of ΔOV with respect to q_1^t.

$\beta_t^{d\tau}$ Probability that total demand of the classes 1 to j-1 during time interval [t,dτ] be more than q_{j-1}^t (Protection level)

$p_t^{d\tau}$ Probability that total demand of the classes j to L during time interval [t,dτ] be more than $q_0^t - q_{j-1}^t$ (Order acceptation limit)

$\bar{P}_{[j,L]}$ Weighted average unit profit of the classes j to L

$\bar{P}_{[1,j-1]}$ Weighted average unit profit of the classes 1 to j-1

$O_{[1,j-1]}^a$ Amount of demand fulfilled for the classes1 to j-1 from the unprotected portion of production capacity

$O_{[j,L]}^a$ Amount of demand fulfilled for the classes j to L from the unprotected portion of production capacity

w Fraction of the order acceptation limit (unprotected capacity) $(q_0^t - q_{j-1}^t)$ which is occupied by the higher priced classes (1 to j-1) in the case that not only demand of lower priced classes is more than the order acceptation limit, but also demand of higher priced classes is over the protection level (q_{j-1}^t)

S_i Desired service level for class i. Service level is defined as the ratio of accepted (fulfilled) orders to arrived orders (demand). It is not rational to define a desired service level for class 1, because when product classes are nested, every order in class 1 will be accepted if enough capacity is available.

π_i Penalty of loss of service level of the class i when the probability of being less than the desired service level is % 100. It is assumed that total penalty has a linear relation with this probability.

- **Note** To determine π_i's, market studies is needed. However, there is a simpler but less accurate method to do that. One may forgo real values of P_i and ask the decision maker to specify relative and comparative significance of accepting orders from each class and failing to achieve desired service level for each other class. This may be carried out through pairwise comparisons like in AHP. Then relative significances of accepting orders can be replaced for P_i's and those of failing to achieve desired service levels for π_i 's.

D_i^t Amount of demand (Number of arrived orders) for class i till t

A_i^t Amount of fulfilled demand (Number of accepted orders) for class i till t

7. REFERENCES

1. Barut, M. and Sridharan, V., "Design and evaluation of a dynamic capacity apportionment procedure", ***European Journal of Operational Research***, Vol. 155, No. 1, (2004), 112–133.
2. Sridharan, V., "Managing capacity in tightly constrained systems", ***Int. J. Production Economics***, Vol. 56-57, No. 1, (1998), 601-610.
3. Smith, B., Leimkuhlcr, J. and Darrow, R., "Yield management at American airlines", ***Interfaces***, Vol. 22, No. 1, (1992), 8-31.
4. Balakrishnan, N., Sridharan, V. and Patterson, J. W., "Rationing capacity between two product classes", ***Decision Sciences***, Vol. 27, No. 2, (1996), 185-214.
5. Harris, F. H. and Pinder, J. P., "A revenue - management approach to demand management and order booking in assemble-to-order manufacturing", ***Journal of Operations Management***, Vol. 13, No. 4, (1995), 299–309.
6. Kimes, S. E., "Yield management: A tool for capacity constrained service firms", ***Journal of Operations Management***, Vol. 8, No. 4, (1989), 348-363.
7. Weatherford, L. R. and Bodily, S. E., "A taxonomy and research overview of perishable-asset revenue management: Yield management, overbooking and pricing", ***Operations Research***, Vol. 40, No. 5, (1992), 831-844.
8. Stevenson, M., Hendry, L.C. and Kingsman, B.G., "A review of production planning and control: The applicability of key concepts to the make-to-order industry", ***International Journal of Production Research***, Vol. 43, No. 5, (2005), 869-898 .
9. Hendry, L.C., Kingsman, B.G. and Cheung, P., "The effect of workload control (WLC) on performance in make-to-order companies", ***Journal of Operations Management***, Vol. 16, No. 1, (1998), 63-75.
10. Sridharan, V. and Balakrishnan, N., "Capacity rationing in multiperiod planning environments", ***Proceedings of the Decision Sciences Institute, Annual Meeting***, Orlando, FL, (1996), 1258-1260.
11. Patterson, J. W., Balakrishnan, N. and Sridharan, V., "An experimental comparison of capacity rationing models", ***International Journal of Production Research***, Vol. 35, No. 6, (1997), 1639-1649.
12. Pak, K. and Piersma, N., "Overview of OR techniques for airline revenue management", ***Statistica Neerlandica***, Vol. 56, No. 4, (2002), 480-496.
13. Fransoo, J. F. and Sridharan, V., "Demand management as a tactical decision tool in capacitated process Industries", Working Paper, Eindhoven University of Technology, Eindhoven, The Netherlands, (1994).
14. Fransoo, J. C., Sridharan, V. and Bertrand, J. "A hierarchical approach for capacity coordination in multiple products single - machine production systems with stationary stochastic demands", ***European Journal of Operational Research***, Vol. 86, No.1, (1995), 57-72.
15. Hopp, W. J. and Sturgies M. L., "Quoting manufacturing due dates subject to a service level constraint", *IIE Transactions*, Vol.32, No.9, (2000), 771-784.
16. Amaro, G., Hendry, L.C. and Kingsman, B.G., "Competitive advantage, customization and a new taxonomy for non make-to-stock companies", ***International Journal of Operations and Production Management***, Vol. 19, No. 4, (1999), 349-371.
17. Pfeifer, P. E., "The airline discount fare allocation problem", *Decision Sciences*, Vol. 20, No. 1, (1989), 149–157.
18. Das, C., "Explicit formulas for the order size and reorder point in certain inventory problems", ***Naval Research Logistics Quarterly***, Vol. 23, No. 1, (1976), 25–30.
19. Hadley, G. and Whitin, T. M., "Analysis of Inventory Systems", Prentice-Hall, Englewood Cliffs, NJ, (1963).