Demand Fulfilment Models for Revenue Management in a Make-to-Stock Production System

Inauguraldissertation
zur Erlangung des akademischen Grades
eines Doktors der Wirtschaftswissenschaften
der Universität Mannheim

vorgelegt von

Yao Yang

November 2014

Dekan: *Dr. Jürgen M. Schneider*

Referent: *Prof. Dr. Moritz Fleischmann*

Korreferent: *Prof. Dr. Herbert Meyr*

Tag der mündlichen Prüfung: 28. November. 2014

Table of Contents

List of Figures

List of Tables

Chapter I

Introduction

1.1 Research Topic and Motivation

After three decades of development, revenue management (RM) has become an active area of research. This concept is not only applied in the traditional service industries such as airlines, hotel and car rental, but also in manufacturing industries. In this thesis, I consider revenue management approaches for demand fulfilment in a make-to-stock (MTS) production system with known exogenous replenishments and stochastic demand from multiple customer classes.

In an MTS system, production is forecast-driven and cannot easily be adjusted to short-term demand fluctuation. Therefore, when demand is higher than supply, it may not be possible to satisfy all incoming customer orders. The manufacturer then has to decide how to allocate the limited supply, i.e. the finished goods inventory, to the customers as different customers may show different profitability or hold different strategic importance. This situation is similar to the traditional airline revenue management problem, in which a fixed number of seats are sold to multiple customer classes. Thus, it is reasonable to expect that demand fulfilment in an MTS system can also benefit from revenue management ideas. The difference is that in an MTS system, the scarce resource to be allocated is the finished goods inventory rather than seats. Unlike flight seats, inventory is storable and can be replenished at certain times. Therefore, inventory holding costs and backlogging costs might be incurred, which makes profit maximization a more appropriate criterion than pure revenue maximization.

Nowadays, in advanced planning systems (APS), the available finished goods inventory is represented by so-called available-to-promise (ATP) quantities, which are derived from mid-term master planning. For demand fulfilment, APS use a two-level planning process to answer real-time customer requests. In the first allocation planning level, customers are segmented based on their profitability and/or strategic importance and the APS then allocate ATP quantities to different delivery periods and customer classes according to certain predetermined allocation rules. In the second order promise level, the allocated ATP (aATP) is consumed by incoming orders based on simple consumption rules such as first-come, first-served (FCFS). The key connection between the two planning levels is that an incoming order can directly consume the aATP quantities that are allocated to its corresponding class. However, if the aATP is not available for the corresponding class, the order promising process has to search for other options to satisfy the order, e.g. by consuming aATP quantities from lower classes if nesting is applied (Kilger & Meyr, 2008).

Clearly, the quality of the allocation rule adopted has a great impact on the performance of demand fulfilment. For example, when supply is scarce, if two customer classes with the same expected demand show very different profitability, it is beneficial to allocate more supply to the more profitable class than giving both classes the same share. In current APS practice, the ATP quantities are normally allocated according to the priority rankings of the customers, the committed forecast, or predetermined split factors, all of which are merely simple heuristic rules and none of which is profit maximizing.

To achieve systematic optimization, researchers have developed different allocation planning approaches. One stream uses deterministic linear programming (DLP) models to maximize the expected profit (Meyr, 2009). The other stream takes a stochastic perspective and models the problem using stochastic dynamic programming (SDP) (Quante, Fleischmann, & Meyr, 2009). Both of these approaches have limitations: the DLP model considers only expected demand and neglects demand uncertainty; therefore, not all information included in the demand distribution is taken into account, which usually makes the solution suboptimal. SDP, however, is computationally expensive and therefore hardly scalable.

The objective of this thesis is thus to develop well-performing and computationally efficient methods to overcome the limitations of the previous approaches. Here, I consider the same problem setting as Quante et al. (2009) and Meyr (2009): an MTS manufacturer is facing stochastic demand from heterogeneous customers with different unit revenues. Inventory replenishments are scheduled exogenously and are deterministic. For each order, the manufacturer decides whether to satisfy it from stock, back-order it at a penalty cost, or reject it in anticipation of more profitable future orders. The objective is to maximize the expected profit over a finite planning horizon, taking into account sales revenues, inventory holding costs and back-order penalties.

I first consider a finite planning horizon in order to make the proposed models comparable to Quante et al.'s (2009) SDP model. Later, I extend all proposed models to rolling horizon planning to bring them closer to real manufacturing practice.

1.2 Chapter Layout and Contributions

In this section, I provide the chapter layout for the remainder of this thesis. In Chapter 2, I first explain the problem setting in detail and set up a common demand fulfilment model for all approaches. Then, the two existing methods, namely Meyr's (2009) DLP model and Quante et al.'s (2009) SDP model are reviewed and I discuss briefly their advantages and shortcomings.

In Chapter 3, based on Meyr's (2009) DLP model, I borrow the safety stock idea from inventory management to account for demand uncertainty. I develop two versions of a safety margin model, which adds safety margins to the relatively more profitable customers. By doing so, I link the traditional inventory/supply chain management world to the emerging revenue management world. To test the performance of the safety margin models systematically, I set up a numerical study test bed using full factorial design. The numerical result shows that by incorporating demand uncertainty, the safety margin models improve the performance of the pure DLP model and perform very close to the SDP model with much less computational effort.

In Chapter 4, to deal with the computational intractability of the SDP model, I consider several approaches to approximate it using the approximate dynamic programming (ADP) algorithm, the basic idea of which is to approximate the value function of the DP using a certain efficient mathematical programming formulation. I consider a deterministic linear programming approximation (Meyr, 2009), a randomized linear programming approximation (Quante, 2008) and an affine functional approximation (Adelman, 2007). As result, I develop three corresponding bid-price control models, namely the DLP-based bid-price control model, the RLP-based bid-price control model and the dynamic bid-price control model. Following the same numerical study framework as in Chapter 3, I analyse the performance of the three proposed bid-price control models. The numerical result shows that the dynamic bid-price control model, as the best-performing method, achieves a close approximation to the optimal SDP model with much lower computational effort.

Without resolving, it provides a better estimation of bid prices and performs substantially better than the other two static models. With frequent resolving, all three models exhibit similar performance. However, due to the fact that frequent resolving is not always realistic in practice, I conclude that the dynamic bid-price control model, which generates close-to-optimal results with tractable computation time, strikes a reasonable balance between performance and computational expense.

In reality, the production process works continuously (unlike in the airline industry) and there is no end to the planning horizon, thus revenue management models for manufacturing should deal with infinite-horizon problems. Therefore, in Chapter 5, I extend all the models to a rolling planning horizon. Based on the numerical study results, I find that the SDP model, although theoretically no longer the optimal ex-ante policy, still outperforms all the other methods proposed. Among all the heuristics, one of the safety margin models provides the closest performance to the SDP model with the least computational effort, which makes it a promising approximation to the SDP and implies its considerable potential for application in real practice.

The thesis concludes in Chapter 6 with a discussion of the results and issues for future research.

Chapter II

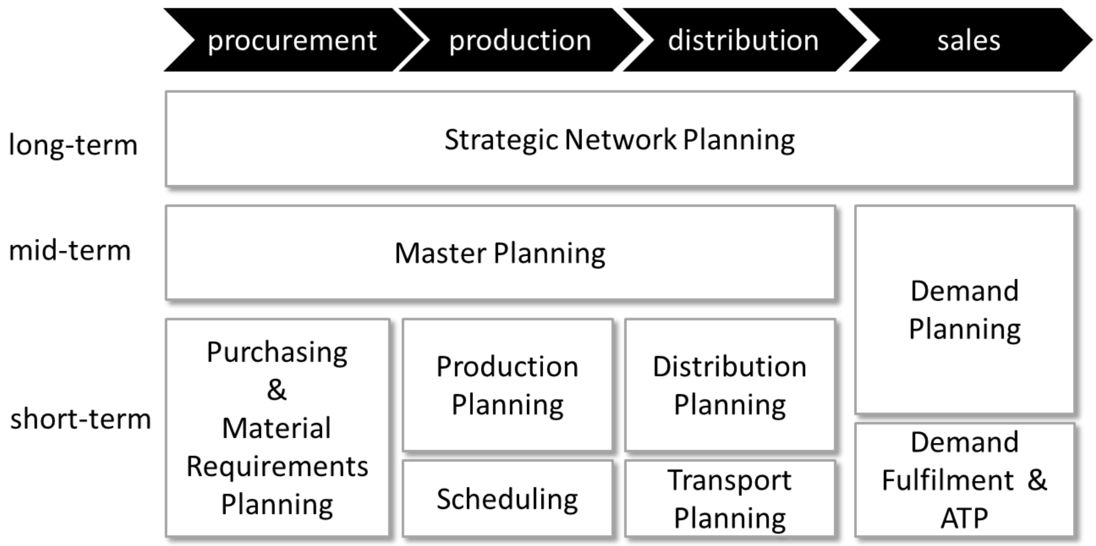
The Demand Fulfilment Model and Previous Research

In this chapter, I first set up a common mathematical model for the MTS demand fulfilment problem considered throughout this thesis. Then, I summarize the two existing approaches from Meyr (2009) and Quante et al. (2009), which serve as a starting point for the work.

2.1 The MTS Demand Fulfilment Problem

As denoted by the following supply chain planning matrix, the component "demand fulfilment & ATP" comprises short-term sales planning, which means fulfilling customer orders based on fixed ATP quantities. This process is similar to the order acceptance problem in traditional airline revenue management. However, in current APS, demand fulfilment solutions are generated based on only simple heuristic rules and no optimization approaches are used. Thus, in this thesis, I use revenue management ideas to optimize the process.

Figure 1 Supply chain planning matrix (Source: Meyr, Wagner, & Rohde, 2008)



I address the same demand fulfilment problem as Meyr (2009) and Quante et al. (2009): I consider a MTS manufacturing system with exogenously determined replenishments and stochastic demand from heterogeneous customers. To maximize the expected profit, the manufacturer has to decide for each arriving order whether to satisfy it from stock, back-order it at a penalty cost, or reject it in anticipation of more profitable future orders. The manufacturer needs to take into account not only sales revenues, but also inventory holding costs and back-order penalties.

I follow the two-level framework of Kilger and Meyr (2008), which comprises an allocation planning level and an order promising level, and summarize the underlying problem description as follows:

(1) There is a finite planning horizon of T, which is subdivided into discrete time periods, $t = 1, \dots, T$.

(2) The inventory replenishment schedule is known and atp_i denotes the ATP quantities arriving at the beginning of period i, $i = 1, \dots, T$.

(3) Customers are differentiated into C different classes, $c = 1, \dots, C$, with corresponding unit revenues of r_c ($r_1 > r_2 > \cdots > r_C$). Orders from different classes arrive in an arbitrary sequence and ask for a random quantity of the products.

(4) It is assumed that the order due dates equal the order arrival date. This assumption is legitimate for the MTS environment as customers normally expect immediate delivery.

(5) D_{ct} denotes the total random demand from Class c with arrival period t. D_{ct} can follow any possible distribution, e.g. Poisson, normal or negative binomial.

(6) At the beginning of the planning horizon, allocation planning is conducted once for the whole planning horizon, with the following information to hand:
   - available inventory that arrives in period i, which is denoted by atp_i;
   - demand forecast: the distribution of D_{ct} is known.

(7) After the allocation planning, incoming orders are processed in real time. Delaying an order causes a back-order cost of b per unit per period and the unit holding cost is h per period.

(8) Partial delivery is allowed.

(9) The objective is to maximize the expected profit, taking into account sales revenues, inventory holding costs and backlogging costs.

Table 1 summarizes the above notations, which are used throughout the thesis.

Table 1 Notations for the demand fulfilment model

Indices:	
$t = 1, \dots, T$	Periods of the planning horizon
$i = 1, \dots, T$	Periods of inventory replenishment
$c = 1, \dots, C$	Customer classes
Data:	
r_c	Unit revenue from customer Class c
b	Unit back-order cost per period
h	Unit holding cost per period
atp_i	Available ATP supply that arrives at the beginning of period i
Random variables:	
D_{ct}	Total demand from Class c with arrival date t, which follows a known distribution with mean μ_{ct} and standard deviation σ_{ct}

2.2 Previous Research

2.2.1 The Stochastic Dynamic Programming (SDP) Model

Quante et al. (2009) model the above demand fulfilment problem using SDP with two additional assumptions: (1) there is at most one order arrival in each period; (2) the demand of a given customer class follows a compound Poisson process and is independent of the demand from other classes and of the available supply.

Using $\vec{x} = (x_1, \dots, x_T)$ as the state variables denoting the available supply quantities and $\vec{u} = (u_1, \dots, u_T)$ as decision variables with u_i denoting the amount of ATP quantities arriving in period i used to satisfy a given order, the additional notations of the SDP model can be summarized as in Table 2.

Table 2 Additional notations for the SDP model

State variables:	
$\vec{x} = (x_1, \dots, x_T)$	Vector of available supply quantities
Decision variables:	
$\vec{u} = (u_1, \dots, u_T)$	Vector of supply quantities used to fulfil a given order
Random variables:	
c	Customer class
d	Order quantity
$F(c, d)$	Joint *cdf* of customer Class *c* and order quantity *d*

(Source: Quante et al., 2009)

Using $V_t(\vec{x})$ to denote the maximum expected profit-to-go from period t to the end of the planning horizon, Quante et al. (2009) develop the following Bellman equation:

$$V_t(\vec{x}) = E_{d,c}\left[\max_{\substack{\vec{u} \\ 0 \le u_i \le x_i, \sum_{i=1}^{T} u_i \le d}} \left\{\sum_{i=1}^{T}(u_i P_t(i,c) - h x_i \delta_{it}) + V_{t+1}(\vec{x} - \vec{u})\right\}\right] \tag{1}$$

where $P_t(i, c)$ is defined as the incremental profit per unit of atp_i used to satisfy one unit of an order of Class c in period t and δ_{it} is defined as 1 if $i \le t$ and 0 otherwise.

After analysing the structural properties, Quante et al. (2009) prove that the optimal policy of the proposed SDP model resembles a booking limit policy, which sets nested protection levels for each class and supply arrival. Supplies are consumed in a first-in-first-out (FIFO) order, i.e. for each incoming order, either the earliest available supply is used to satisfy it or the order is rejected.

In the numerical study, Quante et al. (2009) show that their model outperforms current common fulfilment policies, such as FCFS and the deterministic optimization model provided by Meyr (2009). However, as mentioned in the introduction, because of its high-dimensional state space, this model has very limited scalability.

2.2.2 The Deterministic Linear Programming (DLP) Model

Using the partitioned allocation of each atp_i to Class c with arrival date t, denoted by y_{ict}, as the decision variable, Meyr (2009) models the allocation planning as a DLP problem as follows:

$$max \sum_{i=1}^{T} \sum_{c=1}^{C} \sum_{t=1}^{T} p_{ict} \cdot y_{ict} \tag{2}$$

subject to:

$$\sum_{i=1}^{T} y_{ict} \leq E(D_{ct}) \qquad \forall c, t \tag{3}$$

$$\sum_{c=1}^{C} \sum_{t=1}^{T} y_{ict} \leq atp_i \qquad \forall i \tag{4}$$

$$y_{ict} \geq 0, \text{integer} \qquad \forall i, c, t \tag{5}$$

Here, p_{ict} represents the profit of using one unit of supply i to satisfy the order from customer Class c with arrival date t and can be calculated as follows:

$$p_{ict} = r_c - b(i-t)(1-\delta_{it}) - h(t-i)\delta_{it} \tag{6}$$

Note that the above formulation charges an inventory holding cost only when supply is allocated; the inventory holding cost for unallocated supply is not considered. Although it would be easy to include the inventory holding cost for unallocated supply in the model, it is omitted here to stay in line with the original model (Meyr, 2009). Whether or not the inventory holding cost for unallocated supply is included does not have any impact on the numerical results in this case because the inventory holding cost is sufficiently low that it is beneficial to allocate the supply to some customers whenever possible.

Based on the optimal partitioned allocation quantities, y_{ict}, a rule-based consumption process is used for the order promising.

The DLP model is efficient to solve, but as only the expected demand is taken into account, the performance is not satisfactory if demand uncertainty is high. Quante et al. (2009) show in the numerical study that for low demand variability, the DLP model is competitive with the SDP model, but when demand variability increases, the performance of the DLP model deteriorates drastically.

Chapter III

The Safety Margin Model

3.1 Introduction

To overcome the limitation of the DLP model, I propose a safety margin model which incorporates the impact of demand uncertainty into the deterministic model. I follow a two-level planning process. In the allocation planning level, I allocate the ATP quantities, not only according to the expected demand as Meyr (2009) does, but also borrowing the "safety stock" idea from inventory management to calculate "safety margins" for higher customer classes and set up corresponding booking limits for the lower classes. By doing so, demand uncertainty can successfully be taken into account. For the order promising level, the orders are quoted according to the predetermined booking limits. In a series of numerical simulations, I compare the performance of the safety margin model to other common fulfilment policies.

In summary, this chapter makes the following contributions to the field:

- It presents a new demand fulfilment model which takes customer demand uncertainty into consideration.
- By considering safety margins analogous to safety stocks, I provide insight into the relationship between the traditional inventory/supply chain management world and the relatively new and emerging revenue management world.
- I compare the relative performance of the safety margin model and other fulfilment policies numerically and show that the safety margin model improves the performance of the DLP model with even lower computational expense.

3.2 Literature review

In general, manufacturing systems can be divided into make-to-order (MTO) systems, assemble-to-order (ATO) systems and make-to-stock (MTS) systems. In the literature, most studies regarding revenue management in manufacturing focus on the MTO system. This is due to the direct analogy between the perishable production capacity in MTO and the perishable flight seats in traditional airline revenue management, which makes most of the airline revenue management approaches directly applicable to this environment. Van Slyke and Young (2000), Defregger and Kuhn (2004, 2007), Spengler and Rehkopf (2005), Barut and Sridharan (2005) and Spengler, Rehkopf, and Volling (2007) propose revenue management approaches for the order acceptance problem in the MTO environment. Harris and Pinder (1995) apply revenue management to an ATO environment. Literature on revenue management in the MTS environment is very limited and I shall focus on it in what follows.

Revenue management and manufacturing have significant methodological differences. Whereas revenue management is usually based on stochastic optimization and uses probability distributions to assess opportunity costs, manufacturing companies rely on APS, which take deterministic mathematical programming as the major tool for different planning tasks (Quante et al., 2009). Due to this methodological divide between revenue management and manufacturing, in the literature there are two main streams of research for applying revenue management to demand fulfilment in MTS manufacturing. The first stream adopts the traditional APS perspective and seeks to incorporate revenue management ideas into deterministic optimization. The second stream takes a full stochastic view and models the problem using SDP. In what follows, I briefly review the literature from both research streams.

For the deterministic stream, Kilger and Meyr (2008) set up a two-step framework, in which demand fulfilment is accomplished through ATP allocation and ATP consumption. Ball et al. (2004) propose a similar push-pull framework for ATP models: push-based ATP models pre-allocate available resources to different customer classes and pull-based ATP models promise the allocated resources in direct response to incoming orders. Following this framework, I first consider the allocation models.

Ball, Chen, and Zhao (2004) develop a deterministic optimization-based model that allocates production capacity and raw materials to demand classes in order to maximize profit. They claim that the model is designed for an MTS environment, but actually it is more appropriate for an ATO environment as both capacity and materials are taken into account.

With the same problem setting as in this study, Meyr (2009) proposes a DLP model for ATP allocation. The DLP model maximizes the overall profit and its optimal solution is used as partitioned quantity reserved for each customer class and each arrival period, based on the different consumption rules used for order promising. A numerical study shows that compared to the rule-based allocation methods, this model can significantly improve the performance of APS if demand forecasting is reliable. This DLP model is computationally efficient and can therefore easily be adapted to the APS. However, the major drawback is that it utilizes only expected demand information but ignores demand uncertainty. To overcome this drawback, the safety margin approach extends the DLP model by adding safety margins to expected demand to account for demand uncertainty.

Quante (2008) incorporates demand uncertainty into the DLP model in another way. He adapts the randomized linear programming (RLP) concept derived from Talluri and van Ryzin (1999) to the MTS setting. The idea is repetitively to solve the DLP, not with the expected demand, but with a realization of the random demand with known distribution. The optimal allocation quantity is estimated by a weighted average of the results over all repetitions. The RLP approach is appealing as it is only slightly more complicated than the DLP method but incorporates distributional information on demand. Furthermore, it also has the flexibility to model various possible demand distributions. However, according to Quante's (2008) numerical study, the RLP model does not show promising results and is often dominated by the DLP model.

After allocation planning, aATP quantities could be consumed in real-time mode or batch mode. Kilger and Meyr (2008) propose using search rules for real-time order promising and suggest searching available aATP quantities along three dimensions: customer class, time and product. In order to improve the rule-based consumption methods which represent current practice, Meyr (2009) formulates the real-time order promising problem as a linear programming (LP) model with the objective of maximizing overall profits. To make it easy for practical implementation, he proposes several consumption rules to mimic the LP search process. For batch mode order promising, Fleischmann and Meyr (2003), Pibernik (2005, 2006) and Jung (2010) propose optimization-based models.

For the stochastic stream, Quante et al. (2009) model the demand fulfilment process in MTS production as a network revenue management problem and formulate an SDP model. Unlike the traditional airline network revenue management problem, in the MTS setting, as products are identical, theoretically any of the available supplies can be used to satisfy any incoming order. Therefore, one has to decide not only whether or not to satisfy an order but also which supply and how much of each supply to use as each supply alternative generates a different profit. It transpires that the optimal policy of SDP is the famous booking limit policy, which is easy to implement. Quante et al. (2009) also show that it outperforms current common fulfilment policies, such as FCFS and the deterministic optimization model developed by Meyr (2009). However, because of the “curse of dimensionality”, it is computationally expensive and therefore not really applicable for real-sized problems. In this chapter, I consider the same problem setting as Quante et al. (2009) and compare the performance of their model to the proposed safety margin model in the numerical study.

To address computational intractability, Bertsimas and Popescu (2003) propose a generic approximate dynamic programming (ADP) algorithm, the basic idea of which is to approximate the value function of the dynamic program using a simpler algorithm, such as LP (Erdelyi & Topaloglu, 2010; Spengler et al., 2007; Talluri & van Ryzin, 1999), affine functional approximation (Adelman, 2007) and Lagrangian relaxation approximation (Kunnumkal & Topaloglu, 2010; Topaloglu, 2009). Most of these studies are within the traditional airline revenue management context, indeed to my knowledge, there is no ADP study for the MTS environment.

In addition to the above-mentioned two main streams, there is a paper by Pibernik and Yadav (2009) that is closely linked to the setting of this research: they also consider an MTS system with stochastic demand. However, rather than pursuing the main target of revenue management – profit maximization – the authors still use the traditional service-level maximization as the objective. In addition to this main distinction, other differences include that the authors limit their analysis to two classes and do not allow backlogging.

3.3 Safety Margins

The basic idea of a safety margin is analogous to the use of safety stock in inventory management, i.e. to reserve more stock than expected demand as a "safety margin" for more profitable customers. I first consider a simple single-period, two-class case in which safety margins can be calculated using Littlewood's rule. Then, I generalize the calculation to a multi-period, multi-class case.

3.3.1 Single period, two-class case

I first consider the problem with $T = 1, C = 2$ and assume that within this single period, the lower class (Class 2) arrives before the higher class (Class 1). The problem then becomes the famous Littlewood problem and can be solved directly using Littlewood's rule. I now illustrate how the solution can be interpreted in terms of safety margins.

As the planning horizon consists of only one period, we assume that there is a single inventory replenishment at the beginning of the period, namely atp_1, and use y_1 and y_2 to denote the allocated ATP quantities for Class 1 and Class 2 respectively. Assume the demand of Class 1 is normally distributed with mean μ_1 and standard deviation σ_1. Then, according to Littlewood's rule:

$$y_1^* = \Phi_1^{-1}\left(1 - \frac{r_2}{r_1}\right) = \mu_1 + z_{1-{r_2}/{r_1}} \cdot \sigma_1 \tag{7}$$

i.e. the optimal protection level for Class 1 is y_1^* and the term $z_{1-{r_2}/{r_1}} \cdot \sigma_1$ can be considered the safety margin for Class 1. For Class 2, the corresponding booking limit is then $\left[atp_1 - (\mu_1 + z_{1-{r_2}/{r_1}} \cdot \sigma_1)\right]^+$.

Similar to the safety stock idea, we add a safety margin for the Class 1 customers in the allocation planning stage to afford them better protection.

Incorporating the safety margin of Class 1 into Meyr's (2009) DLP model, which is discussed in the previous chapter, the allocation planning problem can then be modelled as follows:

$$max \quad r_1\, y_1 + r_2 y_2 \tag{8}$$

subject to:

$$y_1 \leq \mu_1 + z_{1-{r_2}/{r_1}} \cdot \sigma_1 \tag{9}$$

$$y_1 + y_2 \leq atp_1 \tag{10}$$

$$y_1, y_2 \geq 0, \text{integer} \tag{11}$$

Constraint (9) modifies the DLP model by adding the safety margin $z_{1-{r_2}/{r_1}} \cdot \sigma_1$ in addition to the mean demand for Class 1. This simple LP forms a continuous knapsack problem the solution to which is equivalent to Littlewood's rule; i.e. by incorporating the safety margin term, we make the DLP model equivalent to the Littlewood model, which is optimal for the single-period, two-class case. This idea can further be extended to the multi-period, multi-class case.

3.3.2 Multi-period, multi-class case

In the demand fulfilment model set out in Chapter 2, the customers are divided into C different classes. In the rest of this chapter, the customers are renamed as K different segments, with $K = C$, as it is necessary to redefine the classes for the multi-period, multi-class case.

Unlike the previous single-period, two-class case, it is difficult to use Littlewood's rule directly to calculate the safety margins for the ATP allocation problem in the MTS setting due to three characteristics. First, it involves multiple customer classes instead of only two. In the MTS setting, there are multiple customer segments and in addition, orders from the same segment with different arrival dates incur different inventory holding or backlogging costs and thus provide different profits. Therefore, these orders cannot be treated as a single class. This cost impact is a major difference between our MTS setting and traditional airline revenue management, where orders from the same customer segment always generate the same profit. Second, the "low-before-high" assumption of Littlewood's rule is violated. The MTS setting involves multiple planning periods and within each period orders from any customer segment may arrive. Therefore, orders that arrive earlier may generate higher profits than orders that arrive later. Third, it considers multiple replenishments, i.e. unlike the single resource case in Littlewood's model, here there are multiple resources to allocate.

In order to deal with the first difficulty mentioned above, i.e. multiple customer classes, I adopt the idea of the expected marginal seat revenue (EMSR) heuristic, which extends Littlewood's rule to the multi-class case (Belobaba, 1989). Thus, each customer segment with a different arrival date is considered as a different class. For a planning horizon of T periods with K customer segments, there are in total $\mathrm{N} = \mathrm{K} \cdot \mathrm{T}$ customer classes.

According to standard EMSR, which also assumes that low-revenue demand arrives before high-revenue demand, the profit ranking of the N classes should correspond to their arrival date, i.e. that with the lowest profit arrives earliest and that with the highest profit arrives latest. With this "low-before-high" assumption, EMSR ensures that the future higher classes are protected against the current lower class. However, this assumption is not sound in the MTS setting as the inherent time structure of the arrival process does not follow the "low-before-high" pattern: each of the N classes has its specified arrival date. Therefore, the second difficulty still remains. In order to address this, as the exact arrival period of each class is known, they are first ranked in descending order of their arrival date. For classes with the same arrival period, their exact arrival sequence is not known and thus we assume that the lower classes arrive before the higher ones, i.e. they are ranked in descending order of their unit revenue, $\mathrm{r_k}$. Then, the first class is the one from Segment 1 that arrives in the last period and the last class is the one from Segment K that arrives in the first period. This ensures that by using EMSR, we are indeed protecting the *future* classes against the current one. Furthermore, at each stage of the EMSR heuristic, when calculating the protection level, only those future classes with a higher profit than the current class are considered. Thus, we also achieve the goal of the standard EMSR, i.e. protecting the *future higher* classes against the current lower class.

To address the third difficulty, namely, the multiple resources, two variants are considered. First, we simply consider the multiple ATP supplies separately, i.e. we calculate the protection levels with respect to each ATP supply as if it were the only resource to allocate without considering the impact of other supplies. The problem with this approach is that it involves "double counting" the demand of the higher classes when calculating protection levels – this method assumes that the future demand can only be fulfilled by a single ATP supply (the one under consideration), whereas in fact it has access to all ATP supplies. One may expect that this "double-counting" problem makes the safety margin model over-protect the higher classes. Therefore, we consider another variant, implicitly allocating the demand to individual supply: for each ATP supply, when determining the corresponding protection levels, we only take the future demand that will arrive before the next supply into account. In contrast to the first case, the potential drawback of this approach is that it may not afford sufficient protection for the higher classes as it considers only a fraction of the demand when calculating the protection levels. The safety margin model adopting the first approach is termed Safety Margin Model_Version 1 (SM_1) and that adopting the second approach is Safety Margin Model_Version 2 (SM_2).

3.3.2.1 Safety Margin Model_Version 1 (SM_1)

Following the two-level planning procedure of APS, SM_1 is first articulated in more detail using the following steps.

Allocation Planning

1. Define classes

   Rank the $N = K \cdot T$ classes in descending order of their due date. Classes with the same due date are ranked in descending order of their unit revenue r_k. Use a new index $j = 1, \dots, N$ to denote customer classes and j can be considered the customer segment/due date combination index. There is a one-to-one correspondence between each j and a combination of k, t.

2. Calculate safety margins

   For each ATP supply, i, do the following calculation:

   a. At stage $j + 1$, let $\mathfrak{I}_{ij}$ denote the set of future classes which have a higher unit profit than class $j + 1$ if atp_i is used, i.e. $\mathfrak{I}_{ij} = \{l \in \{j, j - 1, \dots, 1\}: p_{il} > p_{i,j+1}\}$.

   b. Define the aggregated demand of set $\mathfrak{I}_{ij}$:

$$S_{ij} = \sum_{l \in \mathfrak{I}_{ij}} D_l \tag{12}$$

c. Define the weighted-average profit of set $\Im_{ij}$:

$$\bar{p}_{ij} = \frac{\sum_{l \in \Im_{ij}} p_{il} E[D_l]}{\sum_{l \in \Im_{ij}} E[D_l]} \tag{13}$$

d. Calculate the safety margins

According to Littlewood's rule, the protection level y_{ij}^* for set $\Im_{ij}$ is

$$y_{ij}^* = F_{ij}^{-1}\left(1 - \frac{p_{i,j+1}}{\bar{p}_{ij}}\right) = \bar{\mu}_{ij} + \Delta_{ij} \tag{14}$$

where $\bar{\mu}_{ij} = \sum_{l \in \Im_{ij}} \mu_l$ and Δ_{ij} stands for the safety margin for set $\Im_{ij}$.

If the demand for each Class j is normally distributed with mean μ_j and variance σ_j^2, we have

$$\Delta_{ij} = z_{ij} \cdot \bar{\sigma}_{ij} \tag{15}$$

where

$$\bar{\sigma}_{ij}^2 = \sum_{l \in \Im_{ij}} \sigma_l^2 \tag{16}$$

$$z_{ij} = \Phi^{-1}\left(1 - \frac{p_{i,j+1}}{\bar{p}_{ij}}\right) \tag{17}$$

3. Incorporate safety margins in the DLP model
Adding the safety margins into the DLP model, the resulting allocation planning model is as follows:

$$\max \sum_{j=1}^{N} \sum_{i=1}^{T} p_{ij} \cdot y_{ij} \tag{18}$$

subject to:

$$\sum_{l \in \Im_{ij}} y_{il} \leq \bar{\mu}_{ij} + \Delta_{ij} \qquad \forall i,j \tag{19}$$

$$\sum_{j=1}^{N} y_{ij} \leq atp_i \qquad \forall i \tag{20}$$

$$y_{ij} \geq 0, \text{integer} \qquad \forall i,j \tag{21}$$

Constraint (19) shows that this model does indeed incorporate safety margins in addition to expected demand for the higher classes.

We can use the solution of the above LP as the allocation result. Note that the above LP can actually be decomposed into single-resource problems, i.e. there can be an individual LP for each supply i. This is because in the safety margin calculation (Step 2), we explicitly consider each supply separately and determine the set of future higher classes ($\Im_{ij}$) with respect to the specific supply i. Therefore, the obtained safety margins in Constraint (19) are for each individual supply i. Furthermore, in the above LP, there is no constraint specifying the relation between different supplies.

However, a more convenient way is to write down the corresponding booking limits directly without solving the LP. We are able to do so because Constraint (19) already implies a booking limit for Class $j+1$, namely:

$$b_{i,j+1} = \left[atp_i - (\bar{\mu}_{ij} + \Delta_{ij})\right]^+ \tag{22}$$

Another advantage of using the booking limits directly is that as it is not necessary to know the exact allocation to each class and the protection level term $\bar{\mu}_{ij} + \Delta_{ij}$ in (22) is independent of the real ATP consumption, in the later order processing stage we only need to update the current atp_i quantities before processing each incoming order. It is not necessary to repeat the allocation planning steps all over again. If we use the solution of the above LP as the allocation result, we need frequent re-solving to adapt the allocation to real consumption.

Order Processing

In the order promise stage, we process the incoming orders in real time. The following procedure is used for processing an order from Class *j* ($j = 1, \dots, N$) with an order quantity of *d*:

1. Update the current atp_i quantities for each supply $i = 1, \dots, T$.
2. Determine the corresponding booking limits $b_{ij}, \forall i$ using (22). Note that this way of calculating the safety margin sets nested booking limits for classes with the same arrival period, i.e. within the same period higher classes always have access to units allocated to the lower classes.
3. Search for ATP supplies to fulfil the orders successively in the order of their arrival. Let u_i denote the amount of ATP quantities from supply i used to satisfy the given order and we have the following steps:

   Start with $i = 1$;

   Set $u_i = \max\left(\min\left(b_{ij}, d - \sum_{k=1}^{i-1} u_k\right), 0\right)$;

   Repeat for $i + 1$.

It should be noted that the safety margins and the protection levels from (14) are independent of atp_i. Therefore, before each order processing, it is only necessary to update the current atp_i quantities to determine the current booking limits. It is not necessary to repeat the allocation planning steps.

In the order processing, we start our search for available ATP quantities from the earliest available ATP supply. This is because we know from Quante et al. (2009) that under certain assumptions, the optimal policy for this MTS demand fulfilment situation is also a booking-limit policy and the optimal solution is obtained through a line search, starting with the earliest available supply. Here, we are mimicking the optimal behaviour in the order-processing level.

3.3.2.2 Safety Margin Model_Version 2 (SM_2)

The only difference between SM_2 and SM_1 is that when calculating the protection level with respect to each ATP supply, SM_2 only considers future demand that arrives before the next ATP supply. Therefore, it follows the same procedure as SM_1 and we only need to modify set $\mathfrak{I}_{ij}$ (Step 2a of the allocation planning level) as follows.

For each ATP supply i, assume the next non-zero ATP replenishment arrives at the beginning of period $i+m, m \in \{1, \dots, T-i\}$. At stage $j+1$, $\mathfrak{I}_{ij} = \{l \in \{j, j-1, \dots, 1\}: p_{il} > p_{i,j+1}, t(l) < i+m\}$. As there is a one-to-one correspondence between each class index and k, t combination, $t(l)$ here denotes the arrival date of Class l.

As mentioned above, before each order processing, it is not necessary for the safety margin models to repeat the allocation planning steps as they adopt the booking-limit policy and the safety margins calculated are independent of real consumption. However, in the allocation planning for the DLP model, the available ATP quantities are explicitly allocated to different classes and therefore frequent re-planning is required to adjust the allocation according to real consumption, otherwise performance might suffer. Because of the above-mentioned difference, the safety margin model proposed here is computationally more efficient than the DLP model. I illustrate this further in the next chapter using run-time analysis.

3.4 Numerical Study

To evaluate the performance of different demand fulfilment models, Quante et al. (2009) set up a numerical study framework, comparing their SDP model to a FCFS strategy as well as the DLP model (Meyr, 2009). Following the same assumptions as Quante et al. (2009), both versions of the safety margin models are added to the numerical study framework.

As in Quante et al. (2009), I consider a finite planning horizon here in order to make the models comparable to the SDP model. However, the safety margin models proposed and the DLP model are also applicable in rolling-horizon planning. Within the finite planning horizon, it is not necessary for the safety margin models or the SDP model to do any re-planning because both methods calculate the booking limits up front and the booking limits obtained are independent of real ATP consumption. The DLP model, on the other hand, allocates the current ATP quantities in the allocation planning stage; therefore, frequent re-planning is necessary to enable the allocation to be adjusted according to real consumption.

In what follows, I compare the performance of the safety margin models with the following fulfilment strategies:

- FCFS: a comparison with this strategy shows the benefit of customer segmentation in the demand fulfilment process. To ensure fairness, I limit this policy to fulfilling customer orders only from stock to avoid excessive back-ordering.
- The DLP model (Meyr, 2009): as explained in the previous sections, this strategy allocates the ATP quantities using a DLP model, followed by a rule-based consumption process. The search starts in each incoming order's own priority class. It first looks for aATP quantities that arrive at the required due date. If the order is not fully satisfied, it searches further for aATP quantities that arrive before the due date and then after the due date. Finally, it repeats the search in lower classes. In the numerical study, the DLP model is recalculated after each order processing to ensure its performance is sound. A comparison with this strategy provides an indication of the benefit of incorporating demand uncertainty in the fulfilment process.
- The SDP model (Quante et al., 2009): in this strategy, the optimal policy is also a booking-limit control. This strategy maximizes the expected profit and therefore generates the optimal ex-ante policy.
- Global optimum (GOP): this strategy optimally allocates ATP quantities to demand ex-post and therefore provides the highest achievable profits. In the numerical study, I use it to normalize the results for comparison.

I follow the same assumptions as Quante et al. (2009) for the demand pattern: the orders of a given customer segment follow a compound Poisson process and the order processes of different segments are mutually independent. I discretize the planning horizon in such a way that one order at most could arrive in a single period and the probability of no order arrival is p_0. This single-order-arrival assumption is made for the SDP model as it is required by the Bellman equation formulation, but it not necessary for the safety margin model. For each given arrival, the order size follows a negative binomial distribution (NBD). This choice makes it possible to analyse the effects of large demand variations. In order to make the order size strictly positive, it is modelled as $1 + NB(\mu - 1, \sigma)$, where μ is the mean and σ is the standard deviation. Modelling the ordering process as a compound Poisson process results twofold variability for the customer demand, i.e. the customer demand variability depends on both the variability of the order size and the arrival probabilities.

Based on the above assumptions, I define a numerical experiment with a test bed containing a wide range of problem instances and use simulation to evaluate the performance of the above mentioned models. In subsection 3.4.1, I define the test beds and in subsection 3.4.2, I analyse the results of the numerical study.

3.4.1 Test bed

The test bed is designed based on a full factorial design with five design factors and six fixed parameters. The planning horizon is fixed to 14 periods with two inventory replenishments in period 1 and period 8. The replenishment quantity is fixed to 50 units each time, i.e. $atp_1 = atp_8 = 50$. Three customer segments are considered with different revenues. The inventory holding cost is fixed at $1 per unit per period. It is assumed that the mean demand of each incoming order is constant and equal to 12 units. I summarize the choices for the design factors and fixed parameters in Table 3. This setup is similar to that of Quante et al. (2009); however, they consider only the first three design factors and assume equal order arrival probabilities and a fixed backlogging cost of $10 per unit per period for all customer segments.

The total number of all possible combinations for these design factors is $3^4 \times 4 = 324$, i.e. there are 324 scenarios. For each scenario, I generate 30 different demand profiles and run the corresponding simulations for every policy. In total, this gives $324 \times 30 = 9720$ instances for each policy in the numerical study. This scenario size ensures that both type I and type II errors in the factorial design are limited to 5%.

I now explain the design factors in detail. The first factor in the factorial design is the coefficient of variation of order size (CV). We fix the mean of the order size to $\mu = 12$, but the actual order size can vary from order to order and the variation is represented by the coefficient of variation of the order size $CV = {}^{\sigma}/_{\mu}$, where σ is the standard deviation of the order size. We choose the same range of CV as Quante et al. (2009) to ensure a reasonable range of variability.

The second factor in the factorial design is customer heterogeneity, which is represented by the revenue vector $\boldsymbol{r} = (r_1, r_2, r_3)$ of the customer segments. The revenue vector $(100,90,80)$ represents low customer heterogeneity, whereas

$(100{,}70{,}40)$ represents high customer heterogeneity. These choices are also identical to those of Quante et al. (2009).

Table 3 Design factors and fixed parameters for the numerical study

Name	**Value**
Fixed parameters	
Planning horizon (T)	14
Arrival periods of replenishments	Period 1, Period 8
Replenishment quantity (S)	50
Number of customer segments (K)	3
Inventory holding cost (h)	1
Mean demand per order (μ)	12
Design factors	
Coefficient of variation of order size (CV)	$\left\{\frac{1}{3}, \frac{5}{6}, \frac{4}{3}, \frac{11}{6}\right\}$
Customer heterogeneity ($\boldsymbol{r}$)	$\{(100{,}90{,}80), (100{,}80{,}60), (100{,}70{,}40)\}$
Supply shortage rate (sr)	$\{40\%, 24\%, 1\%\}$
Customer arrival ratio (w)	$\{(1{:}\,2{:}\,3), (1{:}\,1{:}\,1), (3{:}\,2{:}\,1)\}$
Backlogging cost proportion (b)	$\{0.05, 0.1, 0.2\}$

The third factor in the factorial design is the supply shortage rate (sr), which reflects the degree of supply scarcity, defined as follows:

$$sr = 1 - \frac{\sum_{i=1}^{T} atp_i}{(1 - p_0) \times \mu \times T}$$

As, in this case, the supply quantity and the mean demand of each order are both fixed, the supply shortage rate (sr) depends solely on the no arrival probability, p_0. A large p_0 corresponds to a low shortage rate and a small $p_0$indicates a high shortage rate. In the factorial design, we vary sr between 1% and 40% by varying p_0 from 0.4 to 0. We choose these levels because as we only consider situations in which supply is scarce, the 1% shortage rate is almost the lowest shortage rate we can use and 40%

corresponds to a no-arrival probability of 0 and is therefore the highest shortage rate we can use. Quante et al. (2009) use the same levels for the shortage situation, but also consider two more levels for oversupply, i.e. sr being negative.

The fourth factor in the factorial design is the customer arrival ratio (w). This factor reflects the fraction of demand from each customer segment. For instance, when the no-arrival probability $p_0 = 0$, a customer arrival ratio $w = (1:2:3)$ corresponds to an arrival probability of $1/6$ for Segment 1, $1/3$ for Segment 2 and $1/2$ for Segment 3.

The fifth factor in the factorial design is the backlogging cost proportion (b). Quante et al. (2009) assume a fixed backlogging cost for all customer segments. I generalize this assumption to allow different backlogging cost for different customer segments, as customers from different segments pay different prices. In the numerical study, it is assumed that the backlogging cost for different customer segment is proportional to the corresponding revenue. When this proportion is small, e.g. $b = 0.05$, the backlogging penalty is low and when this proportion is large, e.g. $b = 0.2$, the backlogging cost takes 20% of the revenue, which makes the penalty high. Considering the holding cost $h = 1$, the chosen levels of the backlogging cost ratio ensure that the resulting service level is within a reasonable range, e.g. if we fix the other parameters at their middle values (i.e. $CV = \frac{13}{12}, \boldsymbol{r} = (100,80,60), sr = 24\%, w = (1:1:1)$), the replenishment schedule achieves an average cycle service level between 56% and 82% for all segments varying b from 0.05 to 0.2.

3.4.2 Analysis of Results

Using the test bed, we obtain the simulated profits of all the 9,720 instances for each of the fulfilment strategies mentioned in the previous section. The average run time for one simulation instance is 1774.56 seconds for the SDP model, 26.45 seconds for the DLP model, 3.63 seconds for SM_1 and 3.47 seconds for SM_2, using a standard PC with a 2.0GHz Intel Core 2 Duo CPU and 2.00GB memory. The run-time data show that the safety margin models are indeed much more efficient than the SDP model and even faster than the DLP model.

By comparing the simulated profits of other strategies to the simulated profits of the GOP model, we obtain the optimality gaps. We then calculate the average optimality gap for the FCFS strategy, the DLP model, the SDP model and both versions of the SM model over (i) all 9,720 test instances and (ii) all subsets in which one of the design factors is fixed to one of its admissible values. The results are shown in Table 4. As well as the average optimality gap (shown in bold), Table 4 also shows the average backlog percentage (first value in parenthesis), the average lost sales percentage (second value in parenthesis) and the ratio between the average service levels of Segment 1 and Segment 3 (third value in parenthesis) of each strategy. As complementary data, the second and third rows of Table 4 show the average backlogging percentage and average lost sale percentage of each customer segment over all instances for each fulfilment model.

From the first row in Table 4, as expected, we see that the SDP model performs best with an average optimality gap of 3.96%, followed by SM_2 and SM_1 with an average optimality gap of 4.57% and 5.45% respectively. On average, the FCFS strategy (with an optimality gap of 7.55%) performs better than the DLP model (with an optimality gap of 8.84%).

Regarding the safety margin model, apparently both versions are considerably better than the DLP/FCFS models and perform much closer to the SDP model. As the safety margin models are developed to overcome the limitations of the DLP model and the SDP model, in what follows I focus on comparing the safety margin models to these two models to illustrate the difference. By comparing the difference between the optimality gaps, we can see that SM_1 covers approximately 70% of the discrepancy between the DLP model and the optimal SDP model and SM_2 covers 87% of the discrepancy. As the SDP model provides the optimal solution to our problem, we compare the decisions (i.e. the backlogging, lost sale and service level behaviour reflected in the bracketed value of Table 4) made in the two safety margin models and the DLP model to those of the SDP model to understand the profit differences.

Regarding lost sales, the SDP model has an average lost-sales rate of 24.39%. In terms of the different customer segments, it has the highest lost-sales rate for Segment 3 and the lowest rate for Segment 1. If we further consider backlogging behaviour, we can see that it backlogs much more for Segments 1 and 2 than for Segment 3. Based on this observation, we may conclude that compared to the other methods, the SDP model achieves a relatively high service level for the more profitable customers by increasing backlogging.

Table 4 Simulation results

Test bed subset	**N**	**Average optimality gap (%)**				
		FCFS	**DLP**	**SDP**	**SM_1**	**SM_2**
All instances	9720	**7.55**(0.00, 25.39, 1.01)	**8.84**(3.49, 28.11, 1.60)	**3.96**(4.34, 24.39, 1.45)	**5.45**(4.50, 26.61, 1.65)	**4.57**(5.37, 24.33, 1.32)
Avg. backlogging (Seg.1, Seg.2, Seg.3)		*(0.00, 0.00, 0.00)*	*(3.09, 3.38, 2.23)*	*(6.07, 4.19, 1.52)*	*(4.76, 4.43, 2.52)*	*(7.52, 5.48, 1.66)*
Avg. lost sales (Seg.1, Seg.2, Seg.3)		*(0.23, 0.24, 0.24)*	*(0.09, 0.23, 0.43)*	*(0.12, 0.19, 0.39)*	*(0.12, 0.19, 0.47)*	*(0.15, 0.19, 0.36)*
CV = 1/3	2430	**6.49**(0.00, 24.73, 1.02)	**4.33**(4.48, 25.59, 1.96)	**2.57**(3.18, 24.58, 1.82)	**4.49**(4.01, 26.18, 1.89)	**3.82**(5.43, 24.22, 1.43)
CV = 5/6	2430	**7.32**(0.00, 25.30, 1.02)	**6.73**(3.58, 27.05, 1.74)	**3.58**(4.22, 24.66, 1.57)	**4.89**(4.44, 26.54, 1.79)	**4.16**(5.60, 24.31, 1.38)
CV = 4/3	2430	**7.58**(0.00, 25.18, 1.03)	**10.64**(2.61, 28.85, 1.51)	**4.60**(4.36, 24.20, 1.33)	**6.15**(4.41, 26.79, 1.55)	**4.95**(4.98, 24.23, 1.28)
CV = 11/6	2430	**9.04**(0.00, 26.37, 0.98)	**14.70**(3.29, 30.93, 1.31)	**5.34**(5.59, 24.12, 1.19)	**6.48**(5.13, 26.94, 1.44)	**5.53**(5.47, 24.57, 1.19)
r = (100,90,80)	3240	**4.48**(0.00, 25.09, 1.02)	**7.70**(3.36, 27.54, 1.60)	**2.32**(4.43, 23.53, 1.28)	**2.81**(5.58, 23.59, 1.21)	**2.86**(6.00, 23.38, 1.11)
r = (100,80,60)	3240	**7.35**(0.00, 25.58, 1.02)	**8.86**(3.52, 28.34, 1.59)	**4.22**(4.37, 24.54, 1.44)	**5.83**(4.30, 26.60, 1.73)	**4.99**(5.56, 24.31, 1.29)
r = (100,70,40)	3240	**11.44**(0.00, 25.52, 1.00)	**10.19**(3.60, 28.44, 1.61)	**5.63**(4.21, 25.10, 1.66)	**8.20**(3.62, 29.65, 2.34)	**6.16**(4.55, 25.31, 1.63)
sr = 1%	3240	**6.26**(0.00, 13.98, 1.00)	**8.03**(3.16, 15.58, 1.17)	**3.35**(4.73, 11.84, 1.09)	**5.06**(4.43, 14.83, 1.28)	**3.45**(4.50, 12.13, 1.10)
sr = 24%	3240	**7.33**(0.00, 24.61, 1.01)	**9.98**(3.91, 28.27, 1.61)	**4.24**(5.13, 23.61, 1.41)	**5.82**(4.87, 26.26, 1.67)	**4.53**(5.96, 23.48, 1.31)
sr = 40%	3240	**8.75**(0.00, 37.59, 1.04)	**8.42**(3.40, 40.46, 2.36)	**4.16**(3.15, 37.72, 2.31)	**5.40**(4.20, 38.76, 2.39)	**5.47**(5.64, 37.38, 1.74)
w = (1:2:3)	3240	**7.77**(0.00, 25.74, 1.06)	**8.69**(3.85, 27.51, 1.46)	**4.21**(4.36, 24.53, 1.38)	**5.82**(4.37, 26.93, 1.49)	**4.79**(5.35, 24.41, 1.26)
w = (1:1:1)	3240	**7.68**(0.00, 25.00, 1.00)	**8.83**(3.37, 27.74, 1.61)	**4.12**(3.94, 24.32, 1.47)	**5.87**(4.08, 26.82, 1.70)	**4.83**(4.92, 24.31, 1.31)
w = (3:2:1)	3240	**7.25**(0.00, 25.46, 0.97)	**8.99**(3.25, 29.08, 1.78)	**3.60**(4.70, 24.32, 1.50)	**4.76**(5.04, 26.10, 1.79)	**4.15**(5.83, 24.29, 1.38)
b = 0.05	3240	**8.11**(0.00, 25.39, 1.01)	**8.58**(3.71, 27.93, 1.60)	**3.62**(5.84, 23.98, 1.47)	**5.14**(6.45, 26.08, 1.67)	**4.23**(7.39, 23.95, 1.35)
b = 0.1	3240	**7.62**(0.00, 25.39, 1.01)	**8.93**(3.55, 28.10, 1.60)	**4.00**(4.47, 24.31, 1.45)	**5.50**(4.57, 26.55, 1.66)	**4.62**(5.53, 24.25, 1.33)
b = 0.2	3240	**6.92**(0.00, 25.39, 1.01)	**9.03**(3.21, 28.29, 1.60)	**4.25**(2.70, 24.87, 1.42)	**5.71**(2.47, 27.22, 1.63)	**4.86**(3.19, 24.80, 1.28)

Compared to the SDP model, the DLP model has a higher average lost-sales rate (28.11%). However, for Segment 1, its lost-sales rate is even lower than the SDP model, but it loses many more customers from Segments 2 and 3. Regarding backlogging, the DLP model backlogs less on average and does not show a clear differentiation between segments. The backlogging rate for both Segments 1 and 2 are lower than in the SDP model, i.e. the DLP model achieves a higher service level for Segment 1 with even less backlogging, but at the cost of losing many more customers from Segments 2 and 3. This provides clear evidence that the DLP model tends to "over-protect" high profit customers. This over-protection problem in DLP has also been identified by previous studies (De Boer, Freling, & Piersma, 2002).

SM_1 results in a lower lost-sales rate (26.61%) than the DLP model. For Segments 1 and 2, its performance is very close to the SDP model, but for Segment 3, it has the highest lost-sales rate among all the methods. This means that SM_1 also has the over-protection problem, presumably due to the double-counting effect discussed in the previous chapter. Regarding backlogging behaviour, SM_1 has a higher backlogging percentage than the DLP model, especially for Segments 1 and 2. Based on the behaviour pattern of the SDP model, we know that this backlogging behaviour is actually favourable and might be the reason that SM_1 has a lower lost-sales rate compared to the DLP model, which ultimately results in a higher average profit.

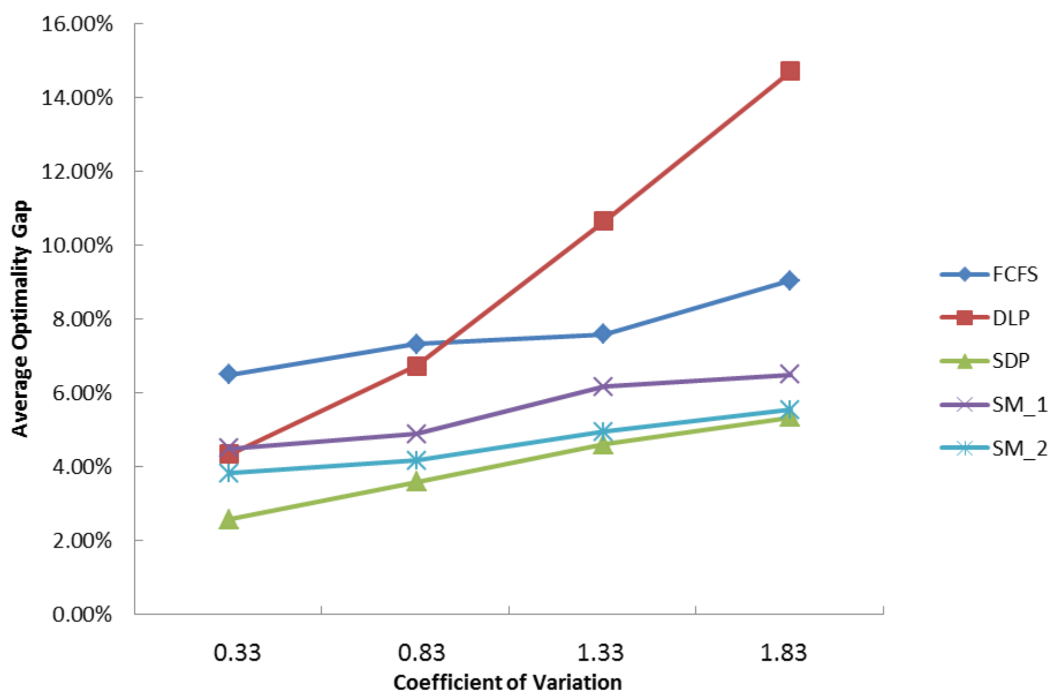
Turning to SM_2, which is proposed to deal with the double-counting effect, from Table 4, we can see that it has the lowest lost-sales rate (24.33%), even lower than the SDP model. This might be because it loses more Segment 1 orders than the other strategies but far fewer Segment 3 orders and therefore does indeed relieve the over-protection problem. Concerning the backlogging behaviour, we can identify that it has the same pattern as the SDP model – increasing backlogging for more profitable customers to achieve a better service level. From Table 4, we can see that SM_2 backlogs even more than the SDP model and this might explain why the average profit of SM_2 is still lower than in the SDP model although it has the lowest lost-sales rate.

The following part of Table 4 provides valuable information on the impact of different design factors on the performance of each fulfilment model. The customer arrival ratio (w) and the backlogging cost proportion (b) have little impact on the performance of the models as for different levels of these two design factors the resulting optimality gaps of each fulfilment model are nearly the same. For the coefficient of variation of order size (CV), customer heterogeneity ($\boldsymbol{r}$) and supply shortage rate (sr), we see that they have a greater impact on the resulting optimality gap of each model and I turn to the analysis of this impact in what follows.

Coefficient of variation of order size (CV)

From Table 4 and the following Figure 2, we can see the clear dependency between the optimality gaps and the CV values.

Figure 2 Average optimality gap for different CV values



Two observations can be made here. (1) In general, as the CV value increases, all strategies show an increasing trend in their average optimality gaps. (2) For small CV values (i.e. low demand variability), the performance of the DLP model and the safety margin models are close to each other. However, as the demand variability increases, the performance of the DLP model drops drastically. On the other hand, the performance of the two safety margin models is always very close to the SDP model and evidently better than the DLP model for larger CV values. As the CV value increases, the gap between SM_2 and the SDP model becomes even closer.

In terms of the first observation, the potential explanation is that the increasing demand variability leads to an increasing forecast error, which harms the performance of every strategy. To explain the other observations regarding the individual performance of each model, I first summarize the response of the SDP model as it provides the “right” response to parameter changes. I then compare the decisions made by the other strategies to this response.

As the CV value increases, SDP is able to keep the average lost-sales rate almost constant. The backlogging percentage increases and the ratio between the average service levels of Segments 1 and 3 decreases. Based on these observations, we may conclude that as demand uncertainty increases, the SDP model reduces the differentiation between segments and backlogs more to retain the average service level.

Regarding backlogging, the response in SM_1 is the same as in the SDP model – it increases the backlogging percentage to cope with the increasing demand uncertainty. It also reduces the differentiation between segments. However, the extent of the reduction is not sufficient as the ratios between the average service levels of Segments 1 and 3 are always higher than that of the SDP model. The above reactions enable SM_1 to keep the lost-sales rate at an almost constant but higher level.

SM_2 does not change the backlogging behaviour too greatly as the CV value increases and the backlogging percentage is kept at a relatively high level. Similar to the SDP model, it also decreases the segment differentiation. The ratios between the average service levels of Segments 1 and 3 are even lower than in the SDP model. The high backlogging percentage and the low segment differentiation enable SM_2 to keep the lost-sales rate as low as in the SDP model, which is ultimately reflected in the very close average profits.

The DLP model fails to retain a constant lost-sales rate. As the CV value increases, the lost-sales rate also increases. Regarding segment differentiation, it responds in the right direction – to reduce the differentiation. But as in SM_1, the extent of the reduction is not sufficient, i.e. it keeps over-protecting the more profitable customers. The DLP model also makes mistakes in the backlogging behaviour: instead of backlogging more to compensate for the increase in uncertainty, it reduces the backlogging percentage as CV increases from 1/3 to 4/3. These mistakes can be attributed to the failure to consider demand uncertainty in the DLP model, resulting in its performance dropping drastically as demand variability increases.

Based on the above analysis, we can conclude that whereas the DLP model fails to provide a satisfactory solution to the problem when demand uncertainty is high, the performance of the safety margin models proposed is promising.

Customer Heterogeneity ($\boldsymbol{r}$)

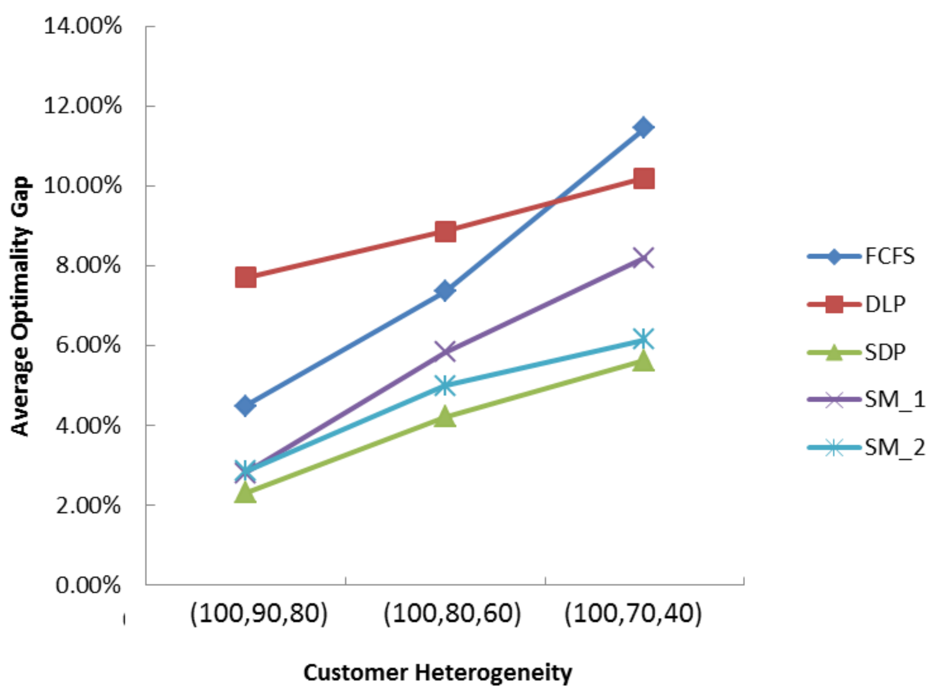
There is also a clear dependency between the resulting average optimality gap and customer heterogeneity. From Table 4 and Figure 3, two key observations can be made. (1) In general, as the scale of customer heterogeneity increases, the performance of all strategies decreases. (2) Although all strategies show the same increasing pattern as the scale of customer heterogeneity increases, the performance difference between strategies is still evident. FCFS is the most affected by increasing heterogeneity, followed by SM_1. On the other hand, the differences between the DLP model, SM_2 and the SDP model are rather constant as heterogeneity increases.

The potential explanation for the first observation might be that when the scale of customer heterogeneity is small, there is no great difference between customer segments. Therefore, the cost of "making mistakes" is low. As the scale of customer heterogeneity increases, the cost of making mistakes also increases, which results in larger optimality gaps.

The main reaction in the SDP model to the increase in customer heterogeneity is to increase the segment differentiation, which is reflected in the increasing value of the ratio between the average service levels of Segment 1 and Segment 3 (third value in parenthesis). This reaction is reasonable because it is more beneficial to ensure better service for the more profitable customers when heterogeneity is high. As segment differentiation increases, the SDP model backlogs less. This is intuitive: from the average backlogging percentage of each segment in Table 4 we know that the SDP model does most of the backlogging for Segments 1 and 2 because it is only cost-effective to backlog the more profitable customers. As segment differentiation increases, the more profitable customers are better protected. Therefore, the need for backlogging decreases. The increasing segment differentiation and the decreasing backlogging percentage lead to an increase in the lost-sales rate.

Both safety margin models react in the same pattern as the SDP model. However, SM_1 tends to overreact to the heterogeneity increase – when heterogeneity is low, the ratio between the average service levels of Segment 1 and Segment 3 is actually small, but the increase in the ratio is much higher than in the SDP model. This might explain why its performance deteriorates when heterogeneity is high. In contrast, the DLP model has a constant average service level ratio, which means it does not react to different heterogeneity levels at all.

Figure 3 Average optimality gap for different customer heterogeneity



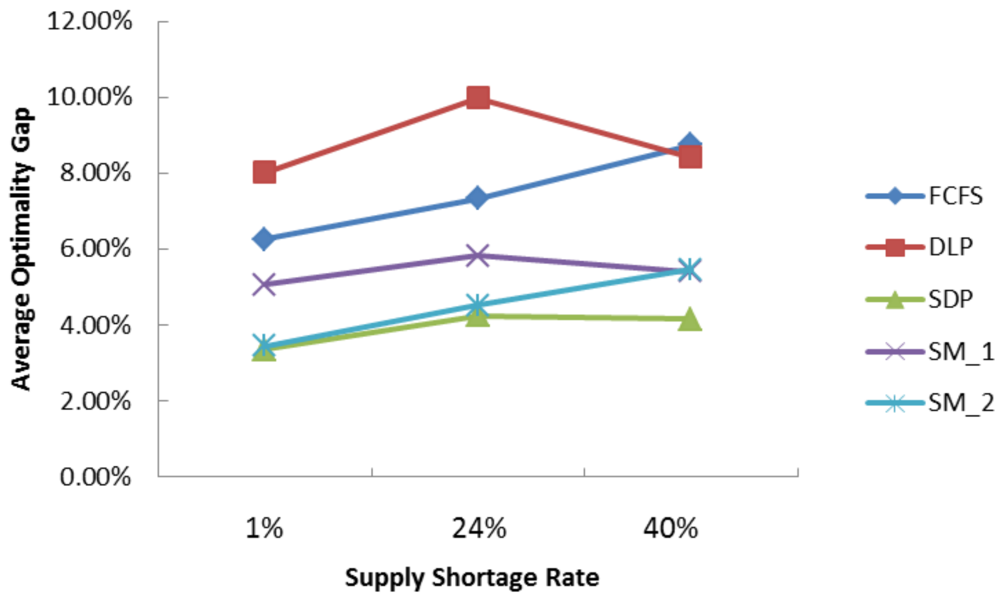
Supply shortage rate (sr)

Finally, I turn to the impact of the degree of supply scarcity. From Table 4 and Figure 4, two observations can be made. (1) The performance of the DLP model, SM_1 and the SDP model shows the same pattern and it is not monotonic in the shortage rate (sr). All strategies perform worst for an intermediate shortage rate of 24%. (2) The performance of SM_2 shows a decreasing pattern as the shortage increases.

In response to increasing shortage, the SDP model increases its segment differentiation. This makes sense as it is beneficial to provide better protection to the more profitable customers when supply is getting scarce. The model's backlogging behaviour is in line with the average optimality gap, which is not monotonic in the shortage rate, and the SDP model backlogs most when the shortage rate is 24%. One reasonable explanation is that for an intermediate shortage rate, resolving the trade-off between selling a unit of supply for current low revenues versus reserving it for future higher revenues is the most difficult. If the level of shortage is very low, the solution is clear and simple: to satisfy all the demand from all segments. If the shortage rate is very high, the solution is also obvious: to reserve enough for the more profitable customers.

The other strategies react in the same way as the SDP model. However, for SM_2, although it also increases segment differentiation as the shortage rate increases, the extent of the increase is not sufficient. When $sr = 1\%$, SM_2 has nearly the same ratio between the average service levels of Segment 1 and Segment 3 as the SDP model. But as the shortage increases, the difference between the ratios becomes larger and larger. When $sr = 40\%$, the average service level ratio of SM_2 is much lower than in the SDP model. This might explain why the performance of SM_2 continues to decrease when the level of shortage increases.

Figure 4 Average optimality gap for different supply scarcity



3.5 Summary

In this chapter, the two-level planning process of the APS has been tracked and two versions of a safety margin model have been developed to allocate the pre-determined ATP quantities to different customer segments with different due date requirements, explicitly taking the demand uncertainty into account by adding safety margins to the relatively more profitable customers.

Based on the DLP model (Meyr, 2009), I borrow the safety stock idea from inventory management to account for demand uncertainty and utilize EMSR to apply it to a multi-class case. By doing so, it is demonstrably possible to link the traditional inventory/supply chain management world successfully to the emerging revenue management world.

The numerical study shows that by incorporating demand uncertainty, the safety margin models do improve the performance of the pure DLP model and provide a close and efficient approximation to the SDP model, which is the optimal ex-ante policy but is computationally very expensive. Therefore, it is possible to conclude that the results highlight the substantial opportunities for improving the demand fulfilment process in MTS manufacturing and that this process could easily be adapted to current APS practice.

The main limitation of the safety margin models is that in the allocation stage, the different supplies are considered separately, which results in the over-protection problem for SM_1 and excessive backlogging for SM_2. Also, there could be other methods for calculating safety margins which might improve performance even further. For the numerical study, a comparison using empirical data instead of theoretical distributions could provide further insight into the relative performance of the different policies.

Chapter IV

Bid-Price Control Models

4.1 Introduction

The main task in the demand fulfilment problem is the same as in the traditional revenue management problem, namely to allocate limited resources to customers with different willingness to pay to maximize revenue or profit. Because this problem considers multiple resources (different replenishments), it is closely linked to the network revenue management problem. However, unlike the traditional network revenue management problem, in which each incoming order requests a specific set of resources, here there is the flexibility to choose between different supply options. This flexibility links the problem in this study to another emerging research topic in the literature, the so-called revenue management problem with flexible products, which can be considered an extension of the traditional network revenue management problem.

Network revenue management is a very important research stream in the revenue management literature as it reflects numerous problems experienced in reality. Generally, it refers to the decision-making problem of selling products that are composed of a bundle of resources under various terms and conditions, with the aim of maximizing revenue (Talluri & van Ryzin, 2004a). In the airline industry, where this class of problem originates, this is mirrored by a network of different flight legs, consisting of a mix of local and connecting traffic. A product is then an “origin-destination itinerary fare class combination”. In the hotel case, each room-night is a separate resource. When customers stay multiple nights, they are consuming multiple resources and the multi-night stays are analogous to multi-leg itineraries in an airline case.

Unlike the single-resource revenue management problem, in the network case, if one of the resources in the bundle faces limitations in its availability, sales of the whole bundle will be constrained. This implies that there are interdependencies between resources and therefore total revenue maximization requires the joint management of all capacity controls across the network (Talluri & van Ryzin, 2004a). In the literature, the interdependencies between resources are sometimes referred to as *network effects*.

In an MTS production system, the resource to be sold is the finished goods inventory. Nowadays, in APS, the available finished goods inventory is represented by the ATP quantities. To satisfy a given order, using ATP quantities from different replenishment batches entails different costs (e.g. inventory holding or backlogging). Therefore, order acceptance in an MTS system also resembles a network revenue management problem, with ATP quantities from different replenishments as different resources. However, unlike a traditional network revenue management problem in which different resources are complementary to each other due to the network effect, in the MTS case, different ATP supplies are substitutive. This is because all finished goods in an MTS setting are physically identical and thus, theoretically, any of the available supplies can be used to satisfy any incoming order and the lack of any specific ATP does not constrain the sale of the others. This flexibility links the problem in this study to the research stream concerning revenue management with flexible products.

The incorporation of flexible products into capacity control is relatively new in revenue management research. A flexible product is defined as a set of alternative products serving the same market (Gallego & Philips, 2004). Purchasers of flexible products are assigned to one of the alternatives at a later time, normally when most of the demand has been realized and uncertainty is lower. Therefore, in revenue management, flexible products are usually provided as supplementary to the more traditional specific products, at a lower price to hedge against demand uncertainty, and they are viewed as inferior to specific products by most customers.

In the MTS setting, ATP quantities from different replenishments can also be treated as different product alternatives, in other words, flexible products. However, there are several differences between the demand fulfilment problem and revenue management with flexible products. First, in an MTS setting, customers do not tend to ask for products from a specific batch. Therefore, there is always the flexibility to choose between different supply alternatives, i.e. there are no specific products. All products in this problem setting are flexible products. Second, the choice of resources has to be made in real time; the order promise cannot be postponed until after most of the demand has been collected. Therefore, the risk-pooling effect of flexible products does not exist. Third, the decision to be made is more complex. In the traditional network revenue management problem, in which only specific products are offered, the decision to be made is a simple “yes” or “no”: whether or not to accept an incoming request. If flexible products are included, one has to go a step further to decide which alternative to assign to an accepted flexible request. In this case, it is necessary to decide not only which alternative to assign, but also how many of each alternative to use as the order size is normally larger than one in an MTS setting. This process can be viewed as repeating the alternative selecting decision multiple times. Fourth, in this problem setting, time plays a particular role in defining the multiple resources. Due to the inventory holding cost and backlogging cost, the margin of choosing a certain resource changes over time. This makes the system more dynamic than the traditional network revenue management case (either with or without flexible products).

From a modelling perspective, theoretically, all network revenue management problems can be modelled using dynamic programming (DP) to determine the optimal policy. The difficulty with DP is that due to the high-dimensional state space, solving the problem analytically usually yields models of intractable complexity, an issue which is commonly referred to as “Bellman’s curse of dimensionality” (Adelman, 2007). Consequently, approximations have been developed that neglect certain factors or estimate certain inputs to generate tractable and implementable solutions, which – despite occasional non-optimality – increase companies’ revenue (Talluri & van Ryzin, 1998).

Of all the methods, bid-price control is becoming the dominant one (Klein & Steinhardt, 2008; Talluri & van Ryzin, 2004a). For network revenue management problems, bid-price control sets a threshold price (bid price) for each resource in the network and an order for a certain product is only accepted if its revenue exceeds the sum of the bid prices of all required resources. From a DP perspective, bid-price control does not in fact always generate the optimal policy for network revenue management problems due to the nonlinearity of the value function. However, it is gaining popularity because of its intuitive nature and the simplicity of implementation.

As discussed in Chapter 2, Quante et al. (2009) also formulate the demand fulfilment problem as an SDP model, the optimal policy of which is a generalization of the booking-limit policy. However, because of the "curse of dimensionality", it is computationally expensive and is therefore not really applicable for real-sized problems.

The purpose of this chapter is then to develop bid-price control methods to solve the demand fulfilment problem in the MTS system. As bid-price controls have proved to be successful in traditional revenue management settings, it is reasonable to expect a similar performance in the MTS environment. However, due to the differences identified between the problem in this study and those in the two research streams mentioned above, it is not just a case of applying the existing methods in a different setting, but also developing bid-price control methods to solve a new and different problem.

In summary, this chapter makes the following contributions to the field:

- It identifies the similarities and differences between the demand fulfilment problem in an MTS system and network revenue management problems.
- Using insights from traditional revenue management settings, I develop three bid-price control models to solve the demand fulfilment problem in an MTS production system.
- I evaluate the performance of the three bid-price control models numerically and compare them to other existing benchmarking methods.

4.2 Literature Review

In the literature, there are different research streams related to solving network revenue management problems. In this section, I only review bid-price control methods. Most of the work on bid-price control in network revenue management problems has taken place within the airline industry and considers only specific products. As an emerging topic, a few papers discuss the situation with flexible products.

4.2.1 Bid-Price Control Model for Network Revenue Management with Specific Products

In general, bid-price control models can be classified as generating a static or dynamic estimate of the marginal value of remaining capacities. While a static model yields bid prices only on basis of the remaining time and capacity at the time of computation, the ultimate goal of dynamic models is to generate bid prices for every possible time-capacity combination until departure (Talluri & van Ryzin, 2004a). Despite the different properties of the bid prices generated, the key ideas behind all of the models are the same: to approximate the DP formulation of the original problem using certain efficient mathematical programming formulations, e.g. LP, and calculate the bid prices by solving the dual problem (Bertsimas & Popescu, 2003).

Static models

Of the models proposed in the literature to compute bid prices, static models are distinguished by the essential characteristic that the resulting bid prices do not change as a function of time or capacity, but stay constant until recomputed.

Williamson (1992) was one of the first to propose DLP to compute bid prices as the optimal dual prices. Assuming demand is equal to its mean, she uses the partitioned allocation of capacity for different products as the decision variable with the objective of maximizing the total revenue. Talluri and van Ryzin (1998) carefully analyse the resulting policy and point out that DLP is actually a linear functional approximation of the DP value function of the network revenue management problem. The main advantage of the DLP model is that it is intuitive and efficient to solve. The weakness is that it treats demand as deterministic and considers only expected demand while neglecting all further distributional information (Kunnumkal & Topaloglu, 2010; Talluri & van Ryzin, 2004a). Despite this shortcoming, several numerical studies have shown that with frequent recalculation, the DLP bid-price control model generates promising performance and outperforms the probabilistic nonlinear programming model (Belobaba, 2001; Belobaba & Lee, 2000; Wiliamson, 1992).

With slight additional complexity, Talluri and van Ryzin (1999) refine the DLP model and incorporate more distributional information in their randomized linear programming (RLP) model by substituting the expected demand with independent samples of the random demand. Talluri and van Ryzin (1999) claim that RLP allows closer to optimal revenue although their computational results do not confirm an absolute dominance over the DLP model in a random network setting. Topaloglu (2009) reinvestigates the relative performance of the DLP and RLP models under different scenarios with different problem parameters and numbers of samples. The results show that on a majority of the test problems, the RLP model is a robust solution method and performs better than the DLP model.

In another attempt to capture the randomness in demand, the probabilistic nonlinear programming (PNLP) method has been developed. Its main difference compared to the DLP model is that PNLP calculates the total revenue based on expected sales instead of the partitioned allocation of capacity, i.e. it considers the possibility that real demand might be lower than the allocated quantity. However, simulations have found that usually it is outperformed by the DLP model (Talluri & van Ryzin, 2004a).

Bertsimas and Popescu (2003) propose an alternative application of a linear approximation to estimate the marginal value of capacities. Instead of computing leg-based bid prices via dual solutions, their certainty equivalent control (CEC) method estimates the opportunity cost for each itinerary by computing the marginal value of capacity. As with typical bid-price controls, a request is accepted if and only if the proposed fare exceeds the estimated opportunity cost. The authors report a revenue increase of 5–10% over the DLP-based bid-price controls. The main disadvantage of the CEC method is that it is necessary to solve a separate LP problem for each product, which is computationally much more expensive than the DLP-based bid-price control method.

Dynamic models

As mentioned above, the static models do not incorporate the dynamics of the underlying system and generate reasonable bid prices only under frequent re-optimization. In practice, however, frequent re-optimization might not be feasible due to the limitations of computational capacity. Thus, a dynamic model, which generates bid prices that vary with time and capacity and therefore can be solved less frequently, is appealing.

To develop such a dynamic model, Adelman (2007) proposes making an affine functional approximation to the value function of the DP model and plugging them into the LP formulation of the DP model. Solving its dual problem with column generation, he obtains a time trajectory of bid prices all at once. In his numerical study, Adelman (2007) shows that the dynamic model outperforms the static bid-price controls by up to 21.4%.

With the same objective of capturing the temporal dynamics of demand, Kunnumkal and Topaloglu (2010) relax the capacity constraints of the DP model using Lagrangian relaxation. Consequently, their method decomposes the optimality equation by periods remaining until departure and yields bid prices that vary with time. The two dynamic models (Adelman, 2007; Kunnumkal & Topaloglu, 2010) generate very similar time trajectories and performance in the proposed settings.

Topaloglu (2009) goes a step further and approaches the network revenue management problem with the goal of computing bid prices that not only encompass the temporal dynamics within the system, but are also contingent on the remaining capacities for the different flight legs. Similar to Kunnumkal and Topaloglu (2010), he uses Lagrangian relaxation to decompose the network revenue management problem into a sequence of single-leg revenue management problems. Concentrating on one flight leg at a time, he generates both capacity- and time-dependent bid prices. Computational experiments indicate that the model outperforms the benchmark strategies such as DLP and RLP and the model proposed by Adelman (2007) within the suggested experimental setup, but with more computational expense.

4.2.2 Network Revenue Management with Flexible Products

As the first publication to introduce the concept of flexible products for revenue management, Gallego and Phillips (2004) consider a simple two-period, two-flight problem for an airline offering a flexible product at a discount in addition to specific products. They provide EMSR-based algorithms for calculating booking limits on both specific and flexible products. The numerical study shows that under reasonable assumptions, offering flexible products generates considerable benefits.

Gallego, Iyengar, Phillips, and Dubey (2004) extend the work of Gallego and Phillips (2004) to a more general network setting with an arbitrary number of products. They consider a continuous time model and approximate the DP model of the resulting network revenue management problem using a DLP method which can be considered a generalization of the LP approximation of the usual network revenue management problem without flexible products, as studied by Williamson (1992) and Talluri and van Ryzin (1998). Using numerical experiments, they verify how the benefits of offering flexible products vary as a function of various parameters, such as time horizon, discount, etc.

With a very similar problem setting to Gallego et al. (2004), Petrick, Steinhardt, Gönsch, and Klein (2012) discretize the planning horizon into individual time periods such that there is at most one order arrival in each period. Unlike Gallego et al. (2004), who assume that the assignment to different alternatives for flexible products can only occur at the end of the planning horizon, Petrick et al. (2012) allow an arbitrary notification date within the planning horizon, during which all flexible requests accepted have to be assigned to an available alternative and after which no more flexible products may be sold. They then provide the DP formulation of the problem and extend three popular static approximation models, namely the DLP, RLP and CEC methods, to the case of flexible products. They report an increase in revenue of up to 4% due to incorporating flexible products and the DLP-based bid-price control model best exploits the additional flexibility.

For the network revenue management problem with or without flexible products, the order acceptance rule is the same: an incoming order is only accepted if there is enough capacity available and its revenue exceeds the sum of the bid prices of all required resources. However, if flexible products are incorporated, this is no longer the end of the story as one still needs to decide which alternative to assign to each accepted flexible request. If this decision is made at the end of the planning horizon, it is possible to achieve an optimal assignment as one has observed all the demands, like Gallego et al. (2004) who develop the assignment problem as an LP model. If an arbitrary notification date is allowed, the problem is more complex as the current assignment can limit the flexibility within the remaining planning horizon. In Petrick et al. (2012), a flexible product is assigned to the alternative with the highest difference between revenue and the corresponding bid prices. Petrick, Gönsch, Steinhardt, and Klein (2010) investigate different assignment mechanisms that differ in the extent to which they exploit the flexibility.

4.2.3 Network Revenue Management in Manufacturing

Literature on bid-price control for the network revenue management problem in manufacturing is very rare. Due to the direct analogy between the perishable production capacity in an MTO system and the perishable flight seats in the airline industry, it is possible to use most of the aforementioned models for the order acceptance problem in an MTO environment. To my knowledge, there is no extant work applying bid-price control in an MTS manufacturing system.

Spengler et al. (2007) implement a static bid-price control to manage the order promising in an MTO system in the iron and steel industry. As the orders obtained in their problem setting are unique and cannot be classified into classes, the standard DLP approximation, which is restricted to multiple fare classes, is not applicable here. Thus, the authors employ a multi-dimensional knapsack problem formulation. According to the computational analysis using real world production data, the proposed bid-price controls perform significantly better than an FCFS strategy.

In terms of the demand fulfilment problem described in Section 1, unlike Quante et al. (2009) who construct it as an SDP model, Meyr (2009) proposes a two-step procedure to solve it: in the first allocation planning step, a DLP model, which is similar to that of Williamson (1992), is developed with the objective of maximizing the overall profit. Its optimal solution is used as partitioned quantity reserved for each customer class and each arrival period. In the second order promising step, allocated quantities are consumed by incoming orders in real time based on certain consumption rules. Following the same framework, Quante (2008) adapts the RLP concept derived from Talluri and van Ryzin (1999) and solves Meyr's (2009) DLP model repetitively with realizations of the random demand with known distribution. The optimal allocation quantity is estimated by a weighted average of the results over all repetitions.

Of all the static models, the DLP (Talluri & van Ryzin, 1998; Williamson, 1992) and the RLP (Talluri & van Ryzin, 1999) are shown to be efficient and perform well. Thus, in this chapter, I use the two models developed by Meyr (2009) and Quante (2008) as the primal problem to calculate the corresponding static bid prices. To capture the temporal dynamics of demand, I adapt Adelman's (2007) affine functional approximation method to calculate the dynamic bid prices.

4.3 Bid-Price Control Models for Demand Fulfilment

As mentioned in the introduction, there are several differences between the order acceptance process in the demand fulfilment problem and that of traditional network revenue management problems. For the traditional case, if only specific products are considered, an incoming order is accepted if and only if its revenue exceeds the sum of the bid prices of the resources required. For the situation in which flexible products are offered, one needs to go a step further to assign an alternative to each flexible request accepted. For the problem in this study, one has to go even further as one still needs to decide how many of each resource to use. Therefore, each of the following bid-price control models proposed contains two steps, namely a bid-price calculation step and an additional order promising step to decide the final consumption scheme.

4.3.1 Bid-price control based on DLP

In the DLP model (equations (2)–(5)), the optimal value of the objective function can be considered an approximation of the value function of the original DP model. Meyr (2009) uses the primal solution directly as the partitioned quantity reserved for each customer class and arrival date, based on which some rule-based order processing methods are used to complete the demand fulfilment problem. Following Williamson (1992) and Talluri and van Ryzin (1998), here we do not use the primal solution of the DLP but calculate the optimal set of dual variables associated with constraint (4) as the bid prices for each corresponding ATP supply.

To process the incoming order, for each supply i, we first calculate the difference between the net profit of using one unit of this supply to satisfy the incoming order and the bid price of this supply:

$$p_{ict} - BP_i \tag{23}$$

Then, we choose the supply with the highest positive difference to satisfy the order. If there is an insufficient quantity in the chosen supply to fulfil the order, we move to the supply with the second highest positive difference and so on, until it is not beneficial to use any supply to satisfy the order, i.e. equation (23) generates only negative results, or there is no more supply available. We then stop the order promising procedure for this order and move to the next one.

To choose supply, we should compare the difference between the profit to be derived from using a certain supply and its corresponding bid price. Here, the sunk inventory holding cost is included in the calculation of the net profit p_{ict}. As p_{ict} is used as the coefficient in the objective function, the resulting bid price BP_i also considers the sunk inventory holding cost. To be consistent with the bid price, we have to use p_{ict} here to calculate the difference as it also includes the sunk inventory holding cost.

4.3.2 Bid-price control based on RLP

Similar to the DLP model in the airline setting, Meyr's (2009) model is efficient to solve, but has been criticized as it neglects demand uncertainty and only takes the expected demand into consideration. To overcome this limitation, Quante (2008) borrows the idea of RLP (Talluri & van Ryzin, 1999) and modifies Meyr's (2009) model by replacing the expected demand in constraint (3) with random demands drawn from the known demand distribution. The resulting LP problem is then solved repetitively, each with an independent sample of the random demand.

In his PhD thesis, Quante (2008) uses the weighted average primal solution as the partitioned allocation quantity. In contrast, here we discard the primal solutions and calculate the RLP-based bid prices based on the associated dual prices. Let us assume that the model is solved N times; it then provides N dual prices for each resource. Following Talluri and van Ryzin (1999), we calculate the final bid price for supply i by taking the average of the N dual prices of supply i.

$$BP_i = \frac{\sum_{n=1}^{N} BP_i^n}{N} \tag{24}$$

where BP_i denotes the final bid price of supply i and BP_i^n is the shadow price of supply i in sample $n(n = 1, \dots, N)$.

The order promising procedure is then the same as the DLP-based bid-price control process in 4.3.1.

4.3.3 Dynamic bid-price control

The above two models generate only static bid prices which do not capture the temporal dynamics of the system. To obtain time-dependent dynamic bid prices, Adelman (2007) derives a model which computes a time trajectory of bid prices all at once. The main steps of this model are as follows: (1) make an affine functional approximation to the value function of the DP model; (2) input the affine functional approximations in the LP formulation of the DP model; (3) solve the dual problem using column generation and obtain the corresponding bid prices. Following these steps, we derive our dynamic bid-price control model in what follows.

We start with the original DP formulation (1). Similar to Adelman (2007), we use the available supply quantities of replenishment i in period t as the basic functions and approximate the value of the state vector $\vec{x}$ using:

$$V_t(\vec{x}) \approx \theta_t + \sum_i V_{t,i} \cdot x_i \qquad \forall t, \vec{x} \tag{25}$$

where the parameter $V_{t,i}$ is the estimation of the marginal value of a unit of supply i in period t, or in other words, the bid price of ATP supply i in period $t-1$, and θ_t is a constant offset. We assume that $V_{T+1,i} = 0$ and $\theta_{T+1} = 0$.

The state vector $\vec{x}$ satisfies

$\vec{x} \in \mathcal{X} \equiv \{\vec{x} \in \mathbb{Z}_+^T : x_i \in \{0,1,\ldots,atp_i\} \ \forall i\}$.

In period 1, we have $\vec{x} = \overrightarrow{atp}$, so for the further analysis, we define

$$\mathcal{X}_t = \begin{cases} \{\overrightarrow{atp}\} & \text{if } t = 1 \\ \mathcal{X} & \text{if } t = 2,\ldots,T \end{cases}$$

Let us assume the maximal possible order size is M; we can then use a $(T \times C \times M)$-vector $\vec{u} \equiv \{u_{i,c,d}\}$ to denote the supply quantity used to satisfy an order from a certain class with a certain order size. When the system is in state $\vec{x}$, this vector has to satisfy

$$\vec{u} \in \mathcal{U}_{\vec{x}} \equiv \{u_{i,c,d} \in \mathbb{Z}_{+}: u_{i,c,d} \leq x_i, \textstyle\sum_i u_{i,c,d} \leq d, \forall i, c, d\} \quad \forall \vec{x}$$

Compared to Adelman (2007), who uses a one-dimensional binary vector to denote the acceptance decision, here we need a three-dimensional integer vector $\vec{u}$. This is because in the traditional airline setting with only specific products, the resources required for a specific order are known, normally defined by an incidence matrix, e.g. matrix $A \equiv (a_{j,k})$, where $a_{j,k} = 1$ if resource j is used by product k and $a_{j,k} = 0$ otherwise (Adelman, 2007; Talluri & van Ryzin, 2004; Topaloglu, 2009). Then, in period t, the decision variable u_t is a binary variable where $u_t = 1$ if the request is accepted and $u_t = 0$ otherwise. In the MTS system, however, the order size is a random variable which is normally larger than one and we do not have the incidence matrix A. Thus, for an incoming order, we have to decide which resource to use and how many of each resource to use as all finished products are physically identical. Therefore, in period t, our decision variable is an integer vector $\vec{u}$. Then, the LP formulation of the DP model from §2.2.1 can be written as follows:

$$(\boldsymbol{D_0}) \quad \min_{V(.)} V_1(\overrightarrow{atp})$$

$$V_t(\vec{x}) \geq \sum_c \sum_d F(c,d) \cdot \left[\sum_i \left(u_{i,c,d} P_t(i,c) - h x_i \delta_{it}\right) + V_{t+1}\left(\vec{x} - \vec{u}_{c,d}\right)\right] + \left(1 - \sum_c \sum_d F(c,d)\right) \cdot V_{t+1}(\vec{x}) \qquad \forall t,\ \vec{x} \in \mathcal{X}_t,\ \vec{u} \in \mathcal{U}_{\vec{x}} \tag{26}$$

Note that as the initial DP formula is different from that of Adelman (2007), the resulting LP formulation is also different here. Substituting the affine functional approximation into the LP formulation, it becomes

$$(\boldsymbol{D_1}) \qquad \min_{\theta, V} \theta_1 + \sum_i V_{1,i} \cdot atp_i \tag{27}$$

$$\theta_t - \theta_{t+1} + \sum_i \left[V_{t,i} \cdot x_i - V_{t+1,i} \left(x_i - \sum_c \sum_d F(c,d) \cdot u_{i,c,d} \right) \right] \geq \sum_c \sum_d F(c,d) \cdot \left[\sum_i \left(u_{i,c,d} P_t(i,c) - h x_i \delta_{it} \right) \right] \quad \forall t,\ \vec{x} \in \mathcal{X}_t,\ \vec{u} \in \mathcal{U}_{\vec{x}} \tag{28}$$

Note that by using equation (25) to approximate the value function, we reduce the number of decision variables from

$1 + (T-1) \cdot \prod_{i=1}^{T} (atp_i + 1)$,

which is exponential in T, to $T(T+1)$. However, $\boldsymbol{D_1}$ still has an exponential number of constraints. Therefore, we use column generation to solve the dual problem $\boldsymbol{P_1}$:

$$(\boldsymbol{P_1}) \qquad z_{p_1} = \max_Y \sum_{t,\vec{x} \in \mathcal{X}_t, \vec{u} \in \mathcal{U}_{\vec{x}}} \left(\sum_c \sum_d F(c,d) \cdot \left[\sum_i \left(u_{i,c,d} P_t(i,c) - h x_i \delta_{it} \right) \right] \right) \cdot Y_{t,\vec{x},\vec{u}} \tag{29}$$

$$\sum_{\vec{x} \in \mathcal{X}_t, \vec{u} \in \mathcal{U}_{\vec{x}}} x_i\, Y_{t,\vec{x},\vec{u}} = \begin{cases} atp_i & \text{if } t = 1, \\ \sum_{\vec{x} \in \mathcal{X}_{t-1}, \vec{u} \in \mathcal{U}_{\vec{x}}} \left(x_i - \sum_c \sum_d F(c,d) \cdot u_{i,c,d} \right) \cdot Y_{t,\vec{x},\vec{u}} & \forall t = 2, \dots, T \end{cases} \quad \forall i,t \tag{30}$$

$$\sum_{\vec{x} \in \mathcal{X}_t, \vec{u} \in \mathcal{U}_{\vec{x}}} Y_{t,\vec{x},\vec{u}} = \begin{cases} 1 & \text{if } t = 1 \\ \sum_{\vec{x} \in \mathcal{X}_{t-1}, \vec{u} \in \mathcal{U}_{\vec{x}}} Y_{t-1,\vec{x},\vec{u}} & \forall t = 2, \dots, T \end{cases} \quad \forall t \tag{31}$$

$Y \geq 0.$

$V_{t,i}$ are then the dual prices on constraint (30) and we can interpret the decision variable $Y_{t,\vec{x},\vec{u}}$ as state-action probabilities as constraint (31) can be rewritten as

$$\sum_{\vec{x} \in \mathcal{X}_t, \vec{u} \in \mathcal{U}_{\vec{x}}} Y_{t,\vec{x},\vec{u}} = 1 \quad \forall t.$$

To use column generation, we first need an initial feasible solution to $\boldsymbol{P_1}$ to start the recursion. Then, we solve $\boldsymbol{P_1}$ with the initial feasible solution to obtain the corresponding dual prices. Using the obtained dual prices as input, we solve the sub-problem to decide whether to add any additional columns to the existing solution set. Finally, we add the chosen columns to the existing solution set and repeat the procedure until the stopping criterion is met.

Similar to Adelman (2007), the "offering nothing" strategy provides a feasible solution to $\boldsymbol{P_1}$, i.e.:

$$\hat{Y}_{t,\vec{x},\vec{u}} = \begin{cases} 1 \text{ if } x_i = atp_i, u_{i,c,d} = 0, \forall i,c,d \\ 0 \qquad\qquad \text{otherwise.} \end{cases} \qquad \forall t, \vec{x} \in \mathcal{X}_t, \vec{u} \in \mathcal{U}_{\vec{x}} \tag{32}$$

Let us assume the resulting dual prices are denoted by V, θ; then, the sub-problem can be written as follows:

$$\max_{t,\vec{x}\in\mathcal{X}_t,\vec{u}\in\mathcal{U}_{\vec{x}}} \pi_{t,\vec{x},\vec{u}} = \max_{t,\vec{x}\in\mathcal{X}_t,\vec{u}\in\mathcal{U}_{\vec{x}}} \sum_c \sum_d F(c,d) \cdot \left[\sum_i \left(u_{i,c,d} P_t(i,c) - h x_i \delta_{it} \right) \right] - \sum_i \left[V_{t,i} \cdot x_i - V_{t+1,i} \left(x_i - \sum_c \sum_d F(c,d) \cdot u_{i,c,d} \right) \right] - \theta_t + \theta_{t+1}$$

which maximizes the reduced profit from (28). When $t = 1$, we have $x_i = atp_i, \forall i$ and for any fixed $t > 1$, the sub-problem can be rewritten as the following integer program, specifying the conditions on the solution set explicitly as constraints:

$$\max_{\vec{x},\vec{u}} \sum_c \sum_d F(c,d) \cdot \left[\sum_i \left(u_{i,c,d} (P_t(i,c) - V_{t+1,i}) - h x_i \delta_{it} \right) \right] - \sum_i \left(V_{t,i} - V_{t+1,i} \right) \cdot x_i - \theta_t + \theta_{t+1} \tag{33}$$

$$u_{i,c,d} \le x_i \qquad \forall i,c,d \tag{34}$$

$$\sum_i u_{i,c,d} \le d \qquad \forall c,d \tag{35}$$

$$x_i \in \{0, \dots, atp_i\} \qquad \forall i \qquad (36)$$

$$u_{i,c,d} \geq 0, \text{integer} \qquad \forall i, c, d \qquad (37)$$

If the objective value of (33) is positive, we add the corresponding column to the existing set of columns for $\boldsymbol{P_1}$, i.e. for each iteration, we do not add only one column (the one with the maximally reduced profit) as the standard column generation algorithm does, but we add a batch of columns, one for each time period, as long as the associated reduced profit is positive.

As a stopping criterion, we specify a percentage, φ, such that as soon as the sum of the optimal objective values of the sub-problems ($\sum_t \pi_t^*$) is smaller than φ per cent of the optimal objective value of $\boldsymbol{P_1}$ with the current set of columns, we stop the column generation iteration.

Using $\mathfrak{J}$ to denote the current set of columns and $Z_{\mathfrak{J}}$ to denote the corresponding optimal objective value of P_1, the column generation algorithm is summarized in Table 5.

Table 5 Column generation algorithm

Algorithm Column generation

Set $\mathfrak{J} = \{(t, \overrightarrow{atp}, \vec{0}) \forall t\}$, solve the restricted problem ($\mathrm{P}_1(\mathfrak{J})$), and set $\pi_t^* = \infty$ for all t.

while $\sum_t \pi_t^* \geq \varphi Z_{\mathfrak{J}}$ do

for all $t \in (1, \dots, T)$

compute $\pi_t^* = \max_{\vec{x},\vec{u}} \pi_{t,\vec{x},\vec{u}}$

select an $(\vec{x}_t, \vec{u}_t) \in \arg\max_{\vec{x},\vec{u}} \pi_{t,\vec{x},\vec{u}}$

update $\mathfrak{J} \leftarrow \mathfrak{J} \cup \{(t, \vec{x}_t, \vec{u}_t)\}$.

solve $\mathrm{P}_1(\mathfrak{J})$)

The order promising procedure is almost the same as that proposed for the DLP-based bid-price control model. The only difference is that for choosing supply, we use the following equation (38) instead of equation (23) to calculate the difference between the profit from using a certain supply and its corresponding bid price: In period t, we compare the incremental profit $P_t(i,c)$ of the incoming order to the *current* bid price of the corresponding supply and calculate the difference:

$$P_t(i,c) - V_{t+1,i} \tag{38}$$

Here, the incremental profit $P_t(i,c)$ is used because the corresponding bid price $V_{t+1,i}$ is calculated based on the profit-to-go, i.e. the sunk inventory holding cost is not included.

4.4 Numerical Study

Following the same numerical study framework as in §3.4.1, this section analyses the performance of the three proposed bid-price control models.

4.4.1 Performance Comparison of Different Bid-Price Control Models

The literature shows that for traditional network revenue management problems, if the static bid-price control models are re-optimized frequently, they perform quite well (Talluri & van Ryzin, 2004a). According to Adelman (2007), resolving the dynamic bid-price control model also leads to a better result. This motivates us to consider the proposed bid-price control models both with and without resolving. The policies considered are summarized as follows:

- **DLP-BPC**: solve the DLP model in §4.3.1 once. Given a fixed set of DLP-based static bid prices, use the order promising procedure in §4.3.1.
- **DLP-BPC Resolved:** resolve the DLP model in §4.3.1 every four periods over the time horizon T. Between solution epochs, use the order promising procedure in §4.3.1.
- **RLP-BPC**: solve the RLP model in §4.3.2 once with $N = 30$. Given a fixed set of RLP-based static bid prices, use the order promising procedure in §4.3.1.
- **RLP-BPC Resolved:** resolve the RLP model in §4.3.2 with $N = 30$ every four periods over the time horizon T. Between solution epochs, use the order promising procedure in §4.3.1.

- **DBPC**: solve the dynamic model in §4.3.3 once. Given a set of fixed dynamic bid prices, use the order promising procedure in §4.3.3.
- **DBPC Resolved**: resolve the dynamic model in §4.3.3 every four periods over the time horizon T. Between solution epochs, use the order promising procedure in §4.3.3.
- **SDP model (SDP):** this strategy applies the optimal policy of the DP formula from Quante et al. (2009) that we are approximating. It provides the optimal ex-ante policy and therefore serves as a benchmark to calculate the optimality gap in the numerical comparison.

For the dynamic bid price control models, we choose the optimality tolerance of $\varphi = 1\%$, i.e. we stop the column generation iteration as soon as the sum of the optimal objective value of the sub-problem is smaller than 1% of the optimal objective value of $\boldsymbol{P_1}$. This optimality tolerance is smaller than Adelman's (2007) 5% and thus provides a more accurate estimation.

Using the test bed, we obtain the simulated profits of all the 9,720 instances for each of the fulfilment strategies. Using a standard PC with a 3.2GHz Intel Core CPU and 32.00GB memory, the average run time for one simulation instance is summarized in Table 6. The run-time data show that all the bid-price control models proposed are much more efficient than the SDP model. The dynamic model takes longer than the static models, but is still tractable.

Table 6 Run-time data

	SDP	DLP-BPC	RLP-BPC	DBPC	DLP-BPC Resolved	RLP-BPC Resolved	DBPC Resolved
Run time (seconds)	1774.56	2.54	3.57	12.35	3.16	3.99	17.82

By comparing the simulated profits of other strategies to the simulated profits of the **SDP** model, we obtain the optimality gaps. We then calculate the average optimality gap for all the above-mentioned models over (i) all 9,720 test instances and (ii) all subsets in which one of the design factors is fixed to one of its admissible values. The results are shown in Table 7. In addition to the average optimality gap (shown in bold), Table 7 also shows the average backlog percentage (first value in parenthesis), the average lost sales percentage (second value in parenthesis) and the ratio between the average service levels of Class 1 and Class 3 (third value in parenthesis) of each strategy. As complementary data, the second and third rows of Table 7 differentiate the average backlogging percentage and average lost sales percentage by customer for each model.

Table 7 Simulation results

Test bed subset	N	Average optimality gap (%)(backlog %, lost sales %, differentiation ratio)			
		SDP	DLP-BPC	RLP-BPC	DBPC
All instances	9720	**0.00**(4.34, 24.39, 1.45)	**7.96**(5.55, 30.55, 1.95)	**6.72**(5.66, 29.59, 1.87)	**3.17**(4.45, 26.90, 1.70)
Avg. backlogging (%) (Cl.1, Cl.2, Cl.3)		*(6.07, 4.19, 1.52)*	*(6.92, 4.92, 3.11)*	*(6.39, 5.43, 3.25)*	*(5.33, 4.16, 1.93)*
Avg. lost sales (%) (Cl.1, Cl.2, Cl.3)		*(0.12, 0.19, 0.39)*	*(0.12, 0.22, 0.55)*	*(0.12, 0.22, 0.53)*	*(0.12, 0.19, 0.49)*
CV = 1/3	2430	**0.00**(3.18, 24.58, 1.82)	**4.59**(4.99, 27.42, 1.93)	**5.91**(4.06, 28.70, 2.10)	**4.68**(2.71, 28.09, 2.23)
CV = 5/6	2430	**0.00**(4.22, 24.66, 1.57)	**7.51**(5.94, 29.92, 1.97)	**8.03**(5.24, 30.72, 2.13)	**3.25**(3.99, 27.40, 1.95)
CV = 4/3	2430	**0.00**(4.36, 24.20, 1.33)	**9.68**(5.35, 31.97, 2.00)	**8.28**(5.68, 30.88, 1.83)	**3.03**(4.61, 26.96, 1.57)
CV = 11/6	2430	**0.00**(5.59, 24.12, 1.19)	**10.76**(5.90, 32.90, 1.91)	**4.50**(7.66, 28.07, 1.51)	**1.40**(6.50, 25.15, 1.30)
r = (100,90,80)	3240	**0.00**(4.43, 23.53, 1.28)	**10.46**(4.11, 31.55, 1.98)	**6.99**(4.91, 28.32, 1.63)	**3.57**(4.13, 25.93, 1.52)
r = (100,80,60)	3240	**0.00**(4.37, 24.54, 1.44)	**7.08**(5.91, 30.48, 1.97)	**6.88**(5.77, 29.89, 1.91)	**2.90**(4.60, 26.84, 1.69)
r = (100,70,40)	3240	**0.00**(4.21, 25.10, 1.66)	**5.85**(6.62, 29.62, 1.91)	**6.21**(6.31, 30.58, 2.12)	**2.98**(4.63, 27.93, 1.95)
sr = 1%	3240	**0.00**(4.73, 11.84, 1.09)	**2.03**(8.63, 10.83, 1.00)	**1.06**(7.62, 11.06, 1.01)	**0.80**(6.25, 11.32, 1.03)
sr = 24%	3240	**0.00**(5.13, 23.61, 1.41)	**8.58**(4.65, 32.75, 2.61)	**4.68**(6.28, 28.42, 2.01)	**2.84**(4.68, 26.77, 1.85)
sr = 40%	3240	**0.00**(3.15, 37.72, 2.31)	**11.98**(3.36, 48.07, 7.03)	**12.98**(3.08, 49.30, 9.24)	**5.32**(2.43, 42.61, 4.25)
w = (1:2:3)	3240	**0.00**(4.36, 24.53, 1.38)	**8.70**(5.78, 30.82, 1.64)	**7.49**(6.00, 29.91, 1.61)	**3.93**(4.29, 27.36, 1.53)
w = (1:1:1)	3240	**0.00**(3.94, 24.32, 1.47)	**7.84**(5.44, 30.27, 2.02)	**5.42**(5.24, 28.97, 2.00)	**3.20**(4.10, 27.06, 1.78)
w = (3:2:1)	3240	**0.00**(4.70, 24.32, 1.50)	**7.45**(5.42, 30.57, 2.32)	**7.27**(5.74, 29.90, 2.06)	**2.51**(4.96, 26.28, 1.84)
b = 0.05	3240	**0.00**(5.84, 23.98, 1.47)	**7.89**(7.84, 30.57, 1.99)	**6.87**(7.67, 29.47, 1.89)	**3.38**(5.68, 26.74, 1.73)
b = 0.1	3240	**0.00**(4.47, 24.31, 1.45)	**7.63**(5.18, 30.33, 1.95)	**6.73**(5.81, 29.43, 1.88)	**3.02**(4.74, 26.59, 1.70)
b = 0.2	3240	**0.00**(2.70, 24.87, 1.42)	**8.37**(3.62, 30.76, 1.91)	**6.54**(3.50, 29.88, 1.83)	**3.13**(2.94, 27.37, 1.68)

Table 7 Simulation results (continued)

Test bed subset	N	Average optimality gap (%)(backlog %, lost sales %, differentiation ratio) DLP-BPC Resolved	RLP-BPC Resolved	DBPC Resolved
All instances	9720	**2.80**(4.75, 25.99, 1.61)	**2.80**(4.63, 26.29, 1.60)	**2.47**(4.97, 26.11, 1.65)
Avg. backlogging (%) (Cl.1, Cl.2, Cl.3)		*(6.56, 4.17, 1.95)*	*(5.99, 4.14, 1.95)*	*(5.95, 4.33, 2.70)*
Avg. lost sales (%) (Cl.1, Cl.2, Cl.3)		*(0.13, 0.21, 0.46)*	*(0.12, 0.22, 0.45)*	*(0.12, 0.19, 0.46)*
CV = 1/3	2430	**1.84**(4.23, 25.36, 1.80)	**2.20**(3.39, 25.99, 1.92)	**2.57**(2.04, 26.81, 2.22)
CV = 5/6	2430	**2.77**(5.00, 26.20, 1.70)	**2.88**(4.49, 26.70, 1.77)	**2.07**(3.42, 26.71, 1.91)
CV = 4/3	2430	**3.20**(4.43, 26.04, 1.53)	**3.28**(4.51, 26.42, 1.47)	**2.39**(4.94, 25.96, 1.47)
CV = 11/6	2430	**3.55**(5.35, 26.35, 1.42)	**2.92**(6.12, 26.07, 1.34)	**2.89**(9.48, 24.96, 1.25)
r = (100,90,80)	3240	**3.32**(3.42, 26.00, 1.58)	**3.08**(3.78, 25.81, 1.49)	**2.74**(4.98, 25.08, 1.46)
r = (100,80,60)	3240	**2.56**(5.10, 26.23, 1.63)	**2.69**(4.74, 26.52, 1.62)	**2.38**(4.98, 26.26, 1.65)
r = (100,70,40)	3240	**2.41**(5.74, 25.73, 1.62)	**2.59**(5.37, 26.55, 1.71)	**2.23**(4.96, 26.99, 1.88)
sr = 1%	3240	**1.97**(6.53, 11.69, 1.07)	**1.18**(5.58, 12.04, 1.09)	**1.02**(5.85, 12.03, 1.09)
sr = 24%	3240	**2.60**(4.66, 25.73, 1.64)	**2.92**(5.23, 25.94, 1.63)	**2.55**(5.36, 26.02, 1.78)
sr = 40%	3240	**3.61**(3.07, 40.54, 3.32)	**3.95**(3.07, 40.90, 3.10)	**3.51**(3.71, 40.27, 3.10)
w = (1:2:3)	3240	**2.89**(5.12, 25.71, 1.39)	**2.85**(5.09, 26.10, 1.44)	**2.61**(4.64, 26.38, 1.53)
w = (1:1:1)	3240	**2.81**(4.65, 25.71, 1.58)	**2.80**(4.18, 26.30, 1.66)	**2.61**(4.64, 26.18, 1.71)
w = (3:2:1)	3240	**2.70**(4.48, 26.54, 1.96)	**2.76**(4.60, 26.49, 1.76)	**2.22**(5.63, 25.77, 1.72)
b = 0.05	3240	**2.59**(6.22, 25.74, 1.63)	**2.71**(6.05, 25.92, 1.63)	**2.64**(6.59, 25.75, 1.66)
b = 0.1	3240	**2.75**(4.68, 26.00, 1.63)	**2.93**(4.75, 26.26, 1.61)	**2.41**(4.99, 26.06, 1.66)
b = 0.2	3240	**3.05**(3.36, 26.22, 1.56)	**2.76**(3.08, 26.70, 1.57)	**2.35**(3.34, 26.52, 1.61)

From the first row in Table 7, we see that without resolving, the performances of the two static bid-price control models are close to each other (with an average optimality of 7.96% for DLP-BPC and 6.72% for RLP-BPC) but are substantially worse than the dynamic model (with an average optimality gap of 3.17%). For all the optimality gaps in Table 7, the 95% confidence intervals are within ± 0.35.

As expected, resolving the bid-price control models improves the performance. The DLP-based bid-price model benefits most from re-optimization with an average optimality gap decrease from 7.96% to 2.80%, while the dynamic bid-price control model benefits least with an average optimality gap decrease from 3.17% to 2.47%. This is intuitive as the DLP-based model takes neither demand uncertainty nor system dynamics into consideration and thus has the highest potential for improvement. On the other hand, the dynamic bid-price model incorporates both factors in the first instance and therefore resolving only leads to marginal improvement.

In fact, with the relatively high resolving frequency of every four periods, the performances of all three bid-price control models are quite similar and are also very close to the dynamic model without resolving. Considering the computational time, the static models with resolving are even more efficient than the dynamic model without resolving. Therefore, one may conclude that for practical purposes it might be better to adopt the static models and resolve them frequently than to use the dynamic model, as the static models generate similar results and are more efficient to solve. However, it must be noted that in practice, very frequent re-optimization is usually not feasible. For instance, in the airline industry re-calculation is normally executed overnight and during the day there is no opportunity for re-optimization. The situation in an MTS production system is similar. In this case, the dynamic model which incorporates system dynamics and generates a bid-price trajectory is much more appealing than the static models which have constant bid prices. This is also the motivation for developing dynamic bid-price control models (Adelman, 2007; Kunnumkal & Topaloglu, 2010; Topaloglu, 2009). The simulation results also show that without frequent resolving, the DBPC model performs much better than the DLP-BPC and RLP-BPC models.

As the SDP model provides the optimal solution to the problem, the decisions (i.e. the backlogging, lost sales and service-level behaviour reflected in the bracketed value of Table 7) made by the bid-price control models are compared to understand their differences in performance. From the first three rows of Table 7, we can see that the three bid-price control models tend to behave quite similarly with frequent resolving: they not only generate a very close average optimality gap, but also have very similar backlogging and lost sales behaviour. Therefore, the resolved versions are treated as one model and the DLP-BPC Resolved model is chosen as representative for the following performance analysis.

Regarding lost sales, the SDP model has an average lost-sales rate of 24.39%. Considering different customer classes, it has the highest lost-sales rate for Class 3 and the lowest rate for Class 1, which shows clear class differentiation. If we further consider its backlogging behaviour, we can see that it backlogs much more for Classes 1 and 2 than for Class 3. This behaviour is reasonable because it is usually more profitable to backlog an order from Class 1 due to its high revenue than to lose it, which leads to a high backlogging rate for Class 1. For Class 3, it is the other way round: it is usually better to keep the supply for future more profitable orders than to backlog it for Class 3.

Compared to the SDP model, the DLP-BPC model has a much higher average lost-sales rate (30.55%). For Class 1, its lost-sales rate is the same as the SDP model, but it loses many more customers from the lower classes; e.g. for Class 3, it loses more than half its customers. Due to the very high lost-sales rate of Class 3, the DLP-BPC model has the highest ratio between the average service levels of Class 1 and Class 3. This shows that the DLP-BPC model tends to over-protect Class 1 customers. Regarding backlogging, the DLP-BPC model backlogs more for each class. This excessive backlogging behaviour suggests that the DLP-BPC model might underestimate the value of the second supply during the demand fulfilment process. I discuss this issue further in the sensitivity analysis.

The RLP-BPC model has a very similar behaviour pattern to the DLP-BPC model but performs slightly better. The average lost-sales rate is 29.55% and it has the same lost-sales rate as the DLP-BPC model for Class 1 and Class 2, but it loses rather fewer Class 3 customers. Regarding backlogging, it backlogs a little more than the DLP-BPC model. Compared to the SDP model, we can see that the RLP-BPC model also has the over-protection problem, but it is less severe than in the DLP-BPC model.

The DBPC model performs closest to the SDP model of all three bid-price control models without resolving. With an average lost-sales rate of 26.90%, it achieves the same service level for Class 1 and Class 2 as the SDP model. For Class 3, its lost-sales rate is higher than in the SDP model but lower than both the DLP-BPC and RLP-BPC models. Regarding backlogging behaviour, its backlogging rate for each class is also lower than both static models, i.e. the DPBC model achieves a higher service level with even less backlogging, which suggests that by incorporating temporal dynamics, the DBPC provides a better estimation of bid prices than the static models.

The DLP-BPC Resolved model performs in quite a similar manner as the DBPC model. It achieves an even lower lost-sales rate of 25.99%. Compared to the static bid-price models without resolving, the DLP-BPC Resolved model loses more Class 1 and Class 2 customers but fewer Class 3 customers, which leads to a lower differentiation ratio. This means that this resolved version relieves the over-protection problem to a certain extent, which contributes to its better performance.

Figure 5 shows the bid-price trajectories of the three models for different shortage rates when the other parameters are fixed to their medium values (i.e. $\mathrm{CV} = \frac{4}{3}, \mathrm{r} = (100{,}80{,}60), \mathrm{w} = (1{:}\,1{:}\,1), \mathrm{b} = 0.1$).

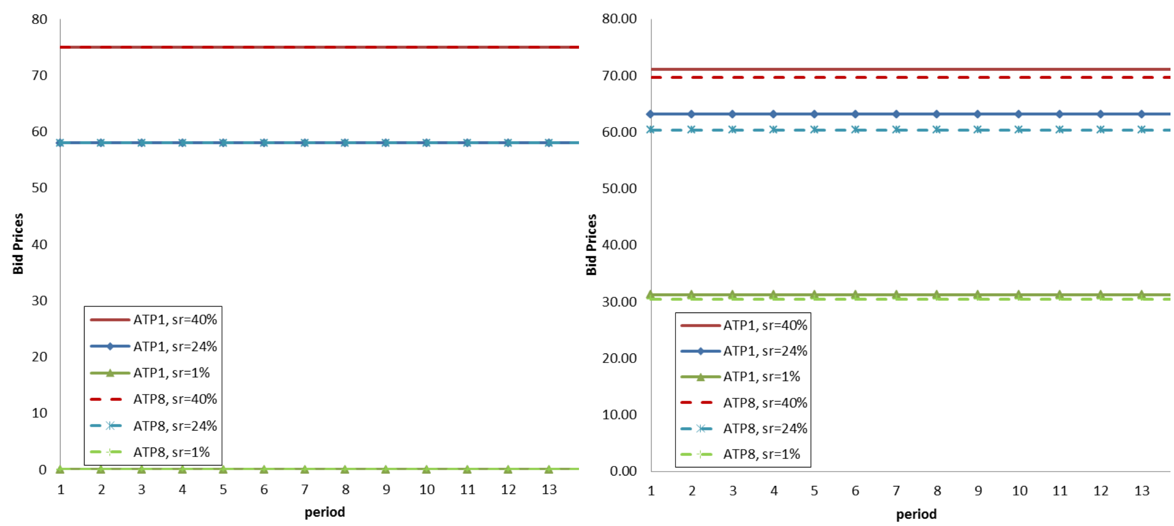
The dynamic bid-price trajectory shows a decreasing pattern in time and its shape is the same as the optimal booking-limit trajectory in Quante et al. (2009). The two curves, representing the bid price of ATP1 and ATP8, converge in period 7, because from period 8 on, the two supplies are the same, i.e. they are both on-hand inventory and generate the same profit for incoming orders. Towards the end of the planning horizon, the bid price drops drastically. This is intuitive as it can be assumed that after the planning horizon, unsold inventory has no value at all.

For the two static models the bid prices are by definition constant and do not change over time. From Figure 5, we can see that when the supply shortage rate is high (sr = 40%), both bid prices generated by the static models are higher than 60, which means Class 3 customers are always rejected. Compared to this, the dynamic model performs more reasonably. Towards the end of the planning horizon, the bid price drops below 60, i.e. Class 3 customers are accepted in the last few periods. This makes sense because at the end of the planning horizon, the chance to sell becomes so slight that one should not miss any incoming orders if one still has inventory on hand. From the above analysis, we can see that to improve performance, updating is necessary for the static models.

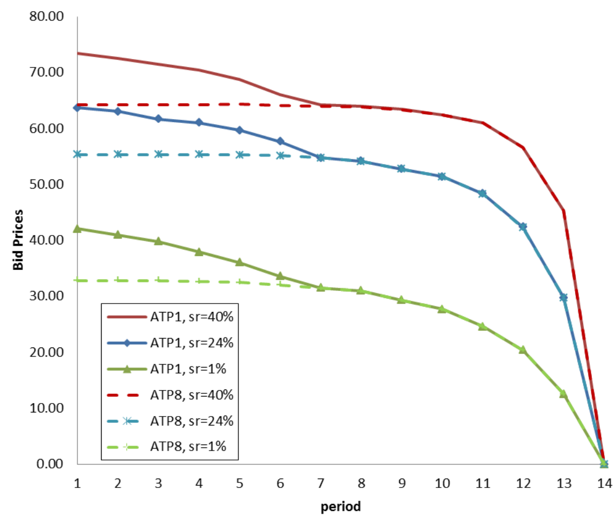
Compared to the RLP-BPC and the DBPC models, the DLP-BPC model tends to overestimate the bid prices when the shortage rate is high and underestimate them when the shortage rate is low. For example, when the supply shortage rate is low (sr = 1%), the bid prices generated by the DLP-BPC model are 0, which makes the DLP-BPC model reduce to an FCFS policy. This might explain its poor performance in Table 7. We also note that in Figure 5(a), the bid price of ATP1 coincides with the bid price of ATP8. This is because in this example backlogging is relatively expensive ($b = 0.1$). The DLP-BPC model tends to avoid any backlogging, which makes the problem in the second supply interval (periods 8–14) a copy of the problem in the first supply interval (periods 1–7). Therefore, the bid prices of the two supplies become the same.

Figure 5 Bid-price trajectories: (a) DLP-BPC, (b) RLP-BPC, (c) DBPC

(a) (b)



(c)



In summary, the following findings are derived from the performance comparison:

- The best-performing method, the DBPC model, achieves a close approximation to the optimal SDP model (with an optimality gap of only 3.17% for the no-resolving version and 2.47% for the resolved version) with much lower computational effort.
- Without resolving, the DBPC model provides a better estimation of bid prices and performs substantially better than the static models.
- The DLP-BPC and RLP-BPC models demonstrate excessive backlogging behaviour, which suggests that they underestimate the value of second supply.
- All bid-price control models tend to over-protect the more profitable customers.
- Resolving improves the performance of the models and the DLP-BPC model benefits the most.
- With resolving, the performance of all three models is very close.

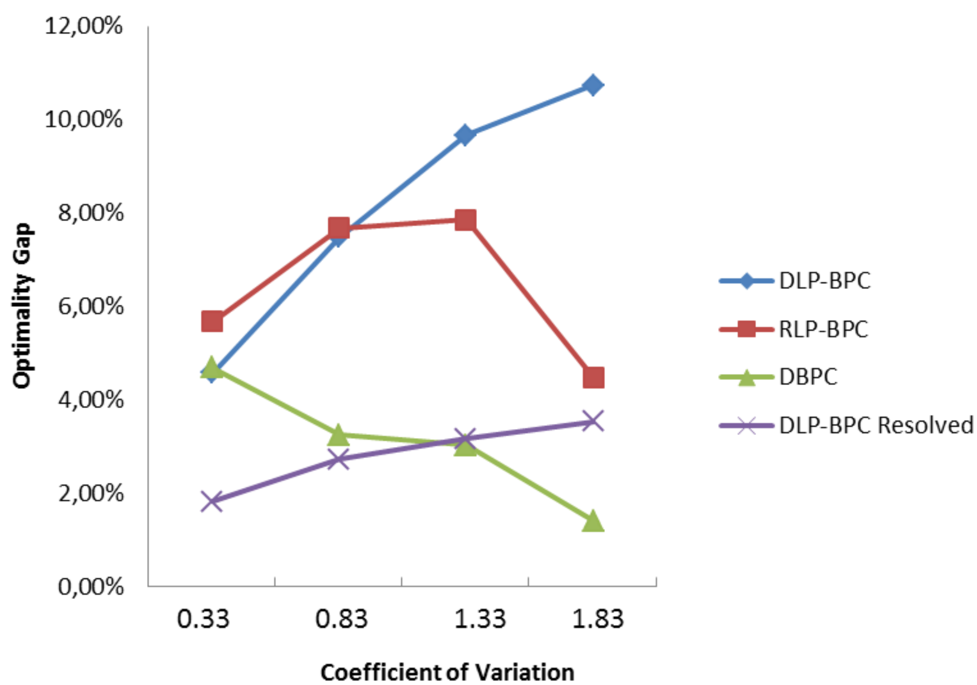
4.4.2 Sensitivity Analysis

The second part of Table 7 provides information concerning the impact of different design factors on the performance of each fulfilment model. The customer arrival ratio (w) and the backlogging cost proportion (b) turn out to have little impact on the performance of the models and thus they are omitted from the sensitivity analysis. The coefficient of variation of the order size (CV), customer heterogeneity (r) and supply shortage rate (sr) have a greater impact on the resulting optimality gap of each model and their impact is discussed in what follows.

Coefficient of variation of order size (CV)

From Table 7 and Figure 6, we can see a clear dependency between the optimality gaps and the CV values.

Figure 6 Average optimality gap for different CV values



From Figure 6, we can observe the following: (1) as the CV value increases, the DLP-BPC model shows a clear increasing trend in its average optimality gap; (2) the RLP-BPC model shows the same trend as CV increases from 0.33 to 1.33, but the optimality gap drops surprisingly as CV increases to 1.83; (3) the DBPC model shows a decreasing pattern in its average optimality gap as demand uncertainty increases. (4) When demand distribution is very low ($CV = 0.33$), the performance of all three bid-price control models (without resolving) are close to each other. As demand variability increases, the dynamic model performs substantially better than the two static models.

(5) The DLP-BPC Resolved model shows same performance pattern as the DLP-BPC model, but performs much better.

The rapidly increasing optimality gap of the DLP-BPC model can potentially be attributed to the fact that this model considers only the expected demand. As the CV value increases, the bid prices generated do not change because the expected demand is constant. Therefore, this model ignores demand uncertainty totally, which makes its lost sales percentage increase and its performance drop drastically as demand variability increases. With resolving, the DLP-BPC Resolved model performs much better because actual demand is incorporated.

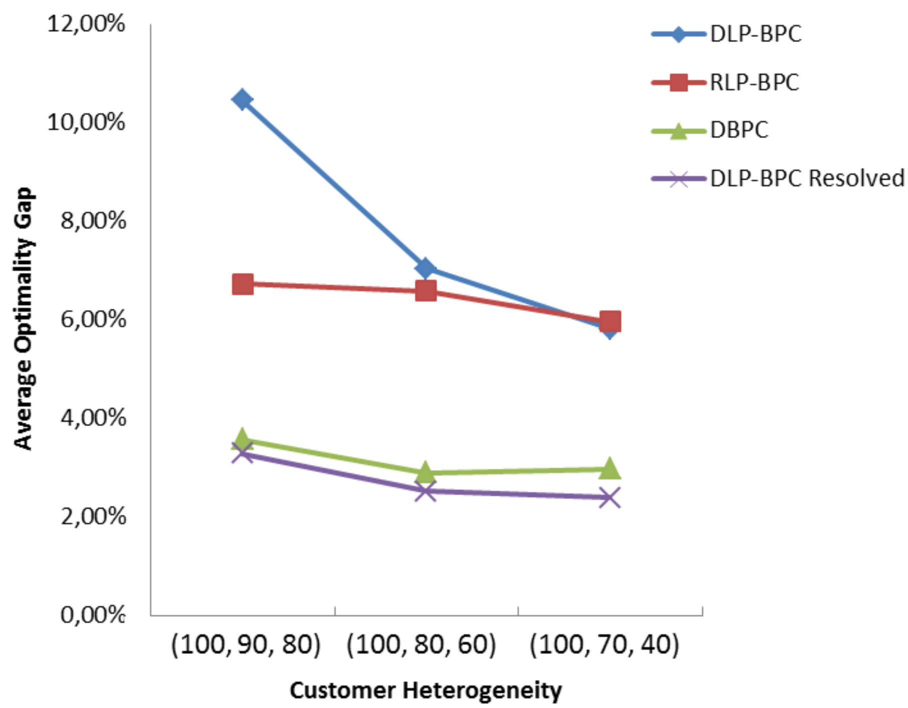
For the RLP-BPC model, when demand uncertainty is low, its performance is close to that of the DLP-BPC model. This is intuitive as when CV is low, the randomly generated demand is close to the mean, which makes the resulting average bid price close to that of the DLP-based version. As CV increases, the RLP-BPC model performs better than the DLP-BPC model and when CV increases to 1.83, its optimality gap even decreases. This might be because when demand uncertainty is high, the randomly generated demand is no longer close to the mean, but represents the real demand distribution to a greater extent. Therefore, the RLP-BPC model generates a better estimation of the bid prices, i.e. the randomization becomes more effective when demand uncertainty is really high.

As CV increases, the DBPC model increases backlogging and reduces class differentiation. By doing so, it reduces the average lost-sales rate as demand uncertainty increases. Therefore, its lost-sales rate becomes increasingly close to that of the SDP model, which might explain the decreasing performance discrepancy between the two models.

<u>Customer heterogeneity (r)</u>

From Table 7 and Figure 7, we can observe that customer heterogeneity shows great impact only on the DLP-BPC model. For the other models, there is no clear dependency between the resulting average optimality gap and customer heterogeneity. For the DLP-PBC model, the optimality gap decreases as the scale of customer heterogeneity increases.

Figure 7 Average optimality gap for different customer heterogeneity



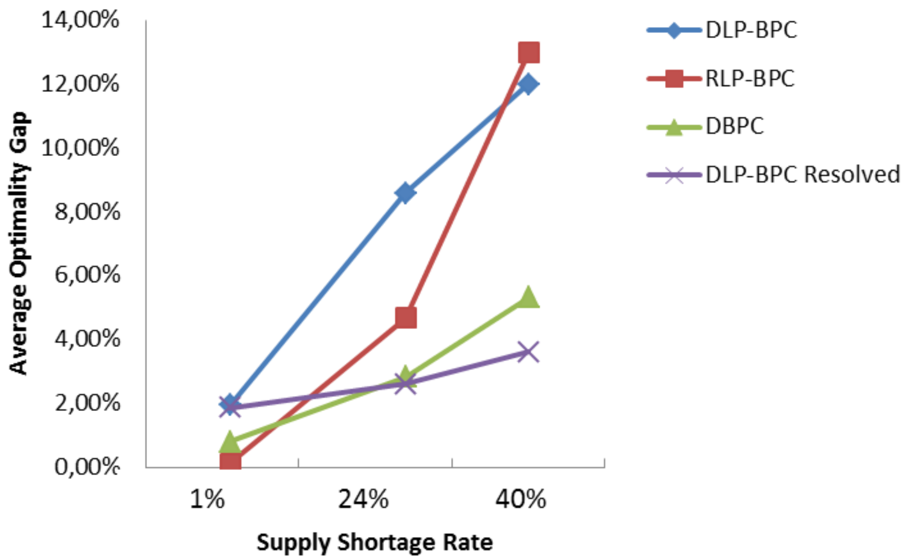
The SDP model's main reaction to an increase in customer heterogeneity is to increase the class differentiation, which is reflected in the increasing value of the ratio between the average service levels of Class 1 and Class 3 (third value in parenthesis). This reaction is reasonable because it is more beneficial to serve the more profitable customers when heterogeneity is high. This increased class differentiation leads to an increase in the lost-sales rate.

However, the DLP-BPC model keeps its differentiation ratio constant, which means it does not react to different heterogeneity levels at all and keeps over-protecting the more profitable customers. From the reaction in the SDP model, we know that this over-protecting behaviour only makes sense when customer heterogeneity is high. Therefore, the optimality gap in the DLP-BPC model decreases as customer heterogeneity increases.

<u>Supply shortage rate (sr)</u>

Finally, we consider the impact of the degree of supply scarcity. From Table 7 and Figure 8, the following is apparent: (1) supply scarcity has a huge impact on the performance of the bid-price control models, especially the two static models; (2) the performance of all three bid-price control models shows a decreasing pattern as the shortage increases.

Figure 8 Average optimality gap for different supply scarcity



The three proposed bid-price control models without resolving and the DLP-BPC Resolved model increase class differentiation as supply becomes scarcer. This is intuitive: as the models aim to keep the same service level for the higher classes, less supply is left for the lower classes when shortage increases.

However, compared to the SDP model, which provides the "right" response to parameter changes, the bid-price control models seem to overreact to a shortage increase – when the shortage rate is low ($sr = 1\%$), the ratios between the average service level of Classes 1 and 3 is actually smaller than in the SDP model, i.e. they do not differentiate enough, but the increase in their ratios is much higher than in the SDP model. For the bid-price control models, when the shortage rate is middling or high ($sr = 24\%, 40\%$), the higher the average service level ratio, the higher the corresponding lost-sales rate and optimality gap, which shows that the bid-price models do indeed overreact to an increase in shortage and therefore their performance is damaged.

Regarding backlogging, all four bid-price models decrease their backlogging behaviour as a shortage increases. This is in line with their differentiation behaviour: as class differentiation increases, the more profitable customers are better served. Therefore, the necessity for backlogging decreases. For the DLP-BPC model, we find that its excessive backlogging mainly happens with a low shortage rate ($sr = 1\%$). From Figure 5, we have already seen that with a low shortage rate ($sr = 1\%$), the DLP-BPC model underestimates the bid price of ATP8 as it is much lower than the estimation of the other two models. But actually from Figure 8 we see that when the shortage rate is low ($sr = 1\%$), the performance of the DLP-BPC model is close to that of the other bid-price control methods. This shows that the excessive backlogging behaviour is not the main reason for the DLP-BPC model's poor performance. The over-protection behaviour, which leads to high lost-sales rate, is the main problem.

4.5 Summary

In this chapter, I have considered the demand fulfilment problem in MTS manufacturing where customers are differentiated into different segments based on their profitability. After discussing the similarities and differences between the demand fulfilment problem and traditional network revenue management problems, three bid-price control models have been developed to solve the problem, based on the idea of approximating the DP formula using simpler mathematical programming.

The numerical study shows that the DBPC model, as the best-performing method, achieves a close approximation to the optimal SDP model but with much lower computational effort. Without frequent resolving, the DBPC model provides a better estimation of bid prices and performs substantially better than the static models.

With resolving, all bid-price control models exhibit similar performance. However, it must be recognized that frequent resolving is usually not feasible in reality. Therefore, the DBPC model, which generates close-to-optimal results with tractable computational time, seems to strike a reasonable balance between performance and computational expense.

Chapter V

Demand Fulfilment Models with a Rolling Planning Horizon

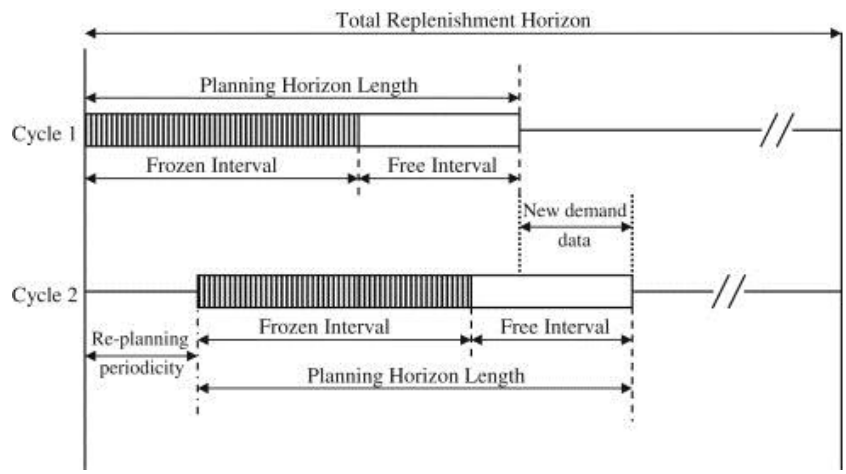
5.1 Introduction

In the previous chapters, the demand fulfilment models have been considered in terms of a finite planning horizon. This is mainly to make the proposed models comparable to the optimal ex-ante SDP model, which assumes a finite planning horizon. However, in practice, unlike the airline industry in which one has a natural end to the planning horizon – the take-off time – production processes are usually on-going without a specific termination time. Therefore, it is reasonable to extend the demand fulfilment models to encompass an infinite planning horizon.

However, modelling and solving infinite horizon planning problems is rather complicated. First, forecasts for the distant future tend to be less precise than for the near future. Therefore, using a very long planning horizon might be of limited use or even counterproductive. Second, the longer the planning horizon, the more information needs to be included, which increases the complexity of the model (Baker, 1977). Thus, for reasons of efficiency and practicality, it is highly desirable to use models that simplify infinite horizon problems and enable decision makers to solve such problems. One common business practice is to solve infinite horizon problems on a rolling horizon basis, creating sequential overlapping finite horizon problems in which only decisions relating to the most immediate periods are implemented before the model is re-run. This process limits dependence on information concerning future events and provides a natural solution to a business environment that entails the on-going nature of activities (Lian, Liu, and Zhu, 2010).

In the literature, rolling horizon planning is used predominantly in production planning. Figure 9 introduces the definition and basic concepts of rolling horizon planning processes.

Figure 9 Illustration of the rolling horizon planning environment (Source: adopted from Narayanan & Robinson, 2010)



In rolling horizon planning systems, a problem of a given temporal length (a finite horizon problem) is solved using information regarding a certain number of future periods (the planning horizon), but only the most immediate decisions are executed. After a pre-specified re-planning period, the system rolls over to the next planning cycle and the latest demand information is applied to update part of the previous schedule which overlaps with the new plan. However, in each subsequent planning cycle, decisions for the frozen interval are not subject to change, but decisions for the free interval may be modified.

To summarize, the planning horizon length (PH) is the number of periods for which the production schedule is developed in each re-planning cycle. The frozen interval (F) covers the scheduled periods within the planning horizon for which decisions are implemented in accordance with the original plan. The re-planning periodicity (RP) is the number of periods between successive re-planning cycles. In a manufacturing resources planning (MRP) system, together with the lot-sizing method used, these three parameters are considered the main policy decision variables that determine the effectiveness of rolling horizon planning systems (Sahin, Naryanan, and Robinson, 2013).

In the production planning context, the initial managerial objective of rolling horizon planning is to satisfy demand at a minimal cost by making the correct production decisions (Sethi & Sorger, 1991). Cost consists of fixed set-up costs and the inventory holding cost (Sahin et al., 2013). Baker (1977) develops a measure called "cost error" which describes the percentage increase in total schedule cost when scheduling takes place on a rolling basis in comparison with an optimal cost that would be achieved if all data were available and known a priori. The second objective is to minimize schedule instability, which is measured by the average changes in the production schedule. A stable schedule is one that does not change with time as new data are added to the planning horizon. General issues related to schedule instability are, among others, its negative effect on workers' willingness to rely on the scheduling system, the higher system costs associated with revising production set-ups and excess inventory (Filho & Fernandes, 2009; Sahin et al., 2013).

In the demand fulfilment problem examined here, schedule instability is not an issue as no schedule is maintained. Therefore, in the next section, I review the literature that examines the impact of the policy decision variables on rolling horizon planning systems, focusing on the cost aspect. As the lot-sizing rule is not of interest, I only concentrate on the other three variables, namely, the planning horizon length (PH), the frozen interval (F) and the re-planning periodicity (RP).

5.2 Literature Review

As mentioned in the previous section, most literature on rolling horizon planning relates to the production planning context. Indeed, I am not aware of any literature that considers rolling horizon planning in a similar problem setting as the demand fulfilment problem examined in this study. In what follows, I carry out a review of a body of literature which is categorized by the three policy decision variables that define the implementation strategies for rolling horizon planning, focusing on their impact on the cost performance of the resulting production schedule.

5.2.1 The impact of planning horizon length

Planning horizon length is also often referred to as the planning window (Quante, 2008). The choice of planning horizon length is very important in designing rolling horizon strategies. Previous studies (Baker, 1977; Bookbinder & H'ng, 1986) have shown that cost error can be limited to 1% when the planning horizon length is chosen properly. It should be noted that the effectiveness of the planning horizon length also depends on the type of demand governing the problem (Zhao & Lam, 1997; Zhao & Lee, 1996).

Assuming deterministic demand, Baker (1977) conducted the first experimental study investigating the effect of planning horizon length on the effectiveness of rolling schedules. This determined that the planning horizon should be at least as long as the natural time between orders (TBO), that is, the order cycle length that would be found using an economic order quantity (EOQ) formula. As the natural order cycle length largely depends on the cost structure of a given problem, the cost structure is identified as a major influence on the optimal planning horizon length (Chung & Krajewski, 1984; Simpson, 1999). Baker (1977) also found that the demand pattern has a significant impact on the effectiveness of a rolling schedule. For a demand pattern without seasonality, the best planning horizon is the natural order cycle (i.e. PH = TBO), while for a demand pattern with a seasonal effect, the optimal planning length is not the natural order cycle, but depends heavily on the seasonal cycle. His finding implies that more information is not always better than less information, which is contradictory to what people usually believe.

Carlson, Beckman, and Kropp (1982) and Blackburn and Millen (1982) elaborate on Baker's (1977) experiment and study the effects of extending the planning horizon under different demand patterns. The results show that an efficient planning horizon is an integer multiple of the natural order cycle, i.e. PH = mTBO (m is a natural number), and extending the planning horizon may increase total cost when the length of the planning horizon is not equal to an integer multiple of the natural order cycle. In their recent numerical study, Narayanan and Robinson (2010) adopted the conclusion of the previous studies (Baker, 1977; Blackburn & Millen, 1982; Carlson et al., 1982) and set the planning horizon length to an integer multiple of the natural order cycle.

There has been little study of the impact of rolling horizon length on stochastic demand. Sridharan and Berry (1990a) conducted an ANOVA analysis and found that increasing the planning horizon under demand uncertainty increases both the schedule cost and instability. Using a simulation technique, Zhao and Lee (1993) show that a planning horizon of PH = 4TBO provides a better solution than PH = 8 TBO for stochastic problems under almost any conditions. A potential explanation might be that under stochastic demand, the decision maker forecasts the demand and uses this forecast to plan. The forecasting accuracy diminishes rapidly the further in the future the information lies as it is less reliable due to demand uncertainty. Another disadvantage of extending the planning horizon is that forecasting information which is further in the future becomes increasingly expensive the longer the horizon and involves a significantly increased computational effort (Bardhan, Dawande, Gavirneni, Mu, & Sethi, 2013).

In summary, previous studies show that under deterministic demand, the planning horizon length should be an integer multiple of the natural order cycle. In the stochastic demand environment, extending the planning horizon may increase total cost.

5.2.2 The impact of a frozen interval

In general, a frozen interval is an interval at the beginning of a planning horizon during which schedules are considered fixed to avoid the negative effects related to excessive schedule changes. By limiting the number of schedule changes, freezing decisions in certain periods can increase schedule stability and limit costs associated with rescheduling. However, freezing too many periods can result in an overall cost increase due to higher changeover and inventory holding costs (Sridharan, Berry, & Udayabhanu, 1987). These costs occur because new information concerning frozen periods is ignored.

In the literature, the frozen interval is normally expressed as a proportion of the planning horizon, i.e. the freezing proportion (Sridharan, Berry, & Udayabhanu, 1988). In practice, decision makers have two ways to determine the freezing proportion, i.e. the period-based and order-based methods. When the period-based method is applied, the freezing proportion is calculated as the number of frozen periods divided by the planning horizon length. In the order-based method, the freezing proportion is calculated as the number of orders frozen divided by the number of order cycles in the planning horizon (Zhao & Lam, 1997). Thus, the freezing proportion has a value ranging from 0.00 to 1.00.

In a deterministic demand environment, earlier studies (Sridharan et al., 1987, 1988; Sridharan & LaForge, 1990) have found that in a single-level system, the cost error ranges from 0.026% when the frozen proportion equals 0.50 to 143.3% when the proportion equals 1.00. That is to say, a small frozen proportion of up to 0.50 has a relatively small effect on costs, whereas a frozen proportion of more than 0.50 results in a substantial cost penalty. This impact on cost is observed to increase rapidly beyond a freezing proportion of 0.80 (Sridharan et al., 1987).

Zhao and Lee (1993, 1996) and Zhao and Lam (1997) consider the impact of the frozen proportion in multi-level systems and come to the conclusion that not all findings derived from the single-level environment can be transferred directly to multi-level problems. Unlike single-level systems, it is more advantageous to freeze a larger proportion in multi-level systems. Here, a higher freezing proportion not only results in lower schedule instability but also in lower schedule costs. Zhao and Lam (1997) recommend a freezing proportion of 0.75 due to its better performance compared to the other freezing proportions tested, i.e. 0.00, 0.25, 0.50 and 1.00. Zhao and Lee (1993) even conclude that freezing the whole planning horizon is often the optimal strategy.

In the stochastic demand environment, studies have reached a consensus that a longer frozen interval results in lower instability and a larger cost error (Sridharan & Berry, 1990a; Sridharan & LaFroge, 1990, 1994; Xie, Zhao, and Lee, 2003). Xie et al. (2003) simulated the impact of freezing proportion under stochastic demand and conclude that if a company wants to reduce total cost, the frozen proportion should be set at 0.00.

5.2.3 The impact of re-planning periodicity (RP)

RP is also often referred to as re-planning frequency and denotes the number of periods between successive re-planning cycles. The greater the re-planning periodicity, the less frequently re-planning occurs and the computational requirement is then reduced. On the other hand, frequent re-planning increases the computational burden but allows the decision maker to consider more reliable, up-to-date data as they become available. To make the best use of the newly available demand information, it is intuitive to make the frozen interval equal to the re-planning interval, i.e. to adopt the new plan immediately once it is made and therefore benefit from the updated demand information. However, due to the stability consideration, the frozen interval in production planning is usually longer than the re-planning interval.

In a deterministic demand environment, Chung and Krajewski (1984) studied the impact of re-planning frequency and found that the product cost structure is important in deciding the appropriate re-planning frequency. If the product cost structure is not extreme, very frequent re-planning is not necessary. Nathan and Venkataraman (1998) found that the length of the planning horizon also has a large impact on the choice of re-planning frequency. A higher re-planning frequency is found to increase total cost exponentially for long planning horizons. Zhao and Lam (1997) observe that as re-planning frequency decreases, both schedule instability and the total cost decrease. This means that less frequent re-planning results in a better overall performance of the production planning system. Moreover, Sridharan and Berry (1990b), Zhao and Lee (1996), Zhao and Lam (1997) and Venkataraman and D'Itri (2001) concur that the best overall performance is achieved by choosing a re-planning frequency equal to the frozen interval, i.e. re-planning takes place once the frozen interval has passed.

Assuming stochastic demand, Lin, Krajewski, Leong, and Benton (1994) carried out a comprehensive study on re-planning frequency. The results show that the choice of re-planning frequency is complex, depending on factors such as cost structure, the length of the planning horizon and the frozen interval, etc. In a single-level system, Sridharan and Berry (1990a) show that the positive impact of a low re-planning frequency increases as the level of demand uncertainty increases. However, in multi-level systems, Sahin et al. (2013) indicate that low re-planning frequency significantly increases costs and instability, making more frequent re-planning preferable. In a case study of a paint company, Nathan and Venkataraman (1998) conclude that more frequent revisions result in higher production and inventory cost. Carlson and Yano (1986) and Yano and Carlson (1985, 1987) also note that frequent re-planning is undesirable under most conditions. They find that it may be more economical to reschedule infrequently and use safety stock to protect against demand uncertainty. In practice, Sahin et al. (2013) observe a tendency for industry planners to re-plan on a weekly basis.

In a nutshell, previous research generally agrees that from a cost perspective, in a deterministic environment, too frequent re-planning may harm performance and that the re-planning interval should equal the frozen interval. Under stochastic demand, in most conditions, the above conclusion still holds.

5.3 Numerical Study using a Rolling Planning Horizon

Thus far, in Chapters 3 and 4, the performance of the different demand fulfilment models has been tested under the assumption of a finite planning horizon. However, an algorithm that performs well using a finite planning horizon does not necessarily provide similar performance in a rolling horizon environment. Therefore, in this section, the performance of the following demand fulfilment models is examined in the case of a rolling planning horizon:

- First-come-first served (FCFS)
- Deterministic linear programming (DLP) model (Meyr, 2009)
- Stochastic dynamic programming (SDP) model (Quante et al., 2009)
- Safety margin model_version 1 (SM_1)
- Safety margin model_version 2 (SM_2)
- DLP-based bid-price control (DLP-BPC)
- RLP-based bid-price control (RLP-BPC)
- Dynamic bid-price control (DBPC)
- Global optimum (GOP)

Similar to the rolling planning horizon system in the production planning context, the rolling horizon approaches are defined as follows:

- There are overlapping planning windows of fixed length, within which all the above models (except GOP) treat the demand fulfilment problem as a finite planning horizon problem and do not do any re-planning.
- During the frozen intervals, for the bid-price control models, the bid prices are fixed; for the two safety margin models, the safety margins (or the protection levels) are fixed; for the DLP and SDP models, the allocated ATP quantities are fixed.

- In the free intervals, after re-planning, it is only necessary to update the corresponding bid prices, safety margins or allocated ATP quantities, which are simply numbers in the APS. No real physical changes are involved. Therefore, as mentioned before, instability is not an issue in the demand fulfilment problem.
- At the end of the total planning horizon, the last re-planning cycle might have a shorter planning window than the previous ones of fixed length as it reaches the end of the total planning horizon. However, as the total planning horizon is rather long compared to a single planning window, this end-of-horizon problem should have little impact on the overall results.

With a finite planning horizon, the SDP model generates the optimal ex-ante solution, but this is not necessarily the case in a rolling horizon environment. Here again, the result derived from GOP is used to normalize the results for comparison. Following the same demand pattern as in Chapter 3, I define a test bed for the numerical experiment in subsection 5.3.1 and analyse the simulation results in subsection 5.3.2.

5.3.1 Test bed

First, it is necessary to define the policy design variables for the rolling horizon planning strategy. According to the literature review, the best planning horizon length is an integer multiple of the natural order cycle for a production planning problem (Baker, 1977; Blackburn & Millen, 1982; Carlson et al., 1982; Narayanan & Robinson, 2010). However, in the demand fulfilment problem considered here, there is no such natural order cycle. Therefore, we first fix the planning horizon length equal to the replenishment cycle, which in this case is the shortest reasonable horizon length. Later, the planning horizon length is extended to a larger integer multiple of the replenishment cycle.

Regarding the frozen interval and re-planning frequency, as the literature indicates (Sridharan & Berry, 1990b; Venkataraman & D’Itri, 2001; Zhao & Lam, 1997; Zhao & Lee, 1996), the best overall performance is achieved by choosing a re-planning frequency equal to the frozen interval. Therefore, in the numerical study, re-planning is always implemented at the end of the frozen interval. We set the freezing proportion at 0.50 to limit its impact on the overall performance (Sridharan et al., 1987, 1988; Sridharan & LaFroge, 1990).

Similar to the numerical study for the finite planning horizon case, we design the test bed based on a full factorial design. Regarding the design factors, in the previous numerical study we find that the customer arrival ratio and the backlogging cost proportion have little influence on the performance of the models. Therefore, only the other three factors are considered here, namely the coefficient of variation of order size, customer heterogeneity and supply shortage rate. Table 8 summarizes the design factors and fixed parameters for the numerical study.

Table 8 Design factors and fixed parameters for the numerical study with a rolling planning horizon

Name	Value
Fixed parameters	
Total simulation horizon (T)	90
Planning horizon length (planning window)	14
Re-planning frequency	7
Replenishment inter-arrival periods	14
Replenishment quantity (S)	100
Number of customer segments (K)	3
Inventory holding cost (h)	1
Backlogging cost proportion (b)	0.1
Customer arrival ratio (w)	$1:1:1$
Mean demand per order (μ)	12
Design factors	
Coefficient of variation of order size (CV)	$\left\{\frac{1}{3},\frac{5}{6},\frac{4}{3},\frac{11}{6}\right\}$
Customer heterogeneity ($\boldsymbol{r}$)	$\{(100,90,80),(100,80,60),(100,70,40)\}$
Supply shortage rate (sr)	$\{40\%, 24\%, 1\%\}$

As Table 8 indicates, we first fix the planning horizon length equal to the replenishment cycle, which is 14 periods in this set-up. Later, this is extended to 28 periods to test its impact on the overall performance.

The total number of all possible combinations for these design factors is $3^2 \times 4 = 36$, i.e. there are 36 scenarios. For each scenario, we again generate 30 different demand profiles and run the corresponding simulations for every policy. In total, this gives $36 \times 30 = 1080$ instances for each policy in the numerical study.

5.3.2 Analysis of Results

5.3.2.1 Performance Comparison of Different Demand Fulfilment Models

Using the test bed, we obtain the simulated profits of all the 1,080 instances for each of the fulfilment strategies. Using a standard PC with a 3.2GHz Intel Core CPU and 32.00GB memory, the average run time for one simulation instance is summarized in Table 9. The run-time data show that in terms of efficiency, the DLP model, the safety margin models and the two static bid-price control models are almost on the same level. The computational effort required by the dynamic bid-price control model is higher by a factor of 40 than the other five models, but this is still much less than the SDP model, the run time of which is higher than the efficient models by a factor of 3000.

Table 9 Run-time data

	DLP	SDP	SM_1	SM_2	DLP_BPC	RLP-BPC	DBPC
Run time (seconds)	2.56	13581.25	4.98	4.74	3.26	4.91	149.18

By comparing the simulated profits of other strategies to the simulated profits of the GOP model, we obtain the optimality gaps. We then calculate the average optimality gap for all the above-mentioned models over (i) all 1,080 test instances and (ii) all subsets in which one of the design factors is fixed to one of its admissible values. The results are shown in Table 10. In addition to the average optimality gap (shown in bold), Table 10 also shows the average backlog percentage (first value in parenthesis), the average lost sales percentage (second value in parenthesis) and the ratio between the average service levels of Class 1 and Class 3 (third value in parenthesis) of each strategy. As complementary data, the second and third rows of Table 10 differentiate the average backlogging percentage and average lost sales percentage by customer for each model.

Table 10 Simulation results

Test bed subset	**N**	**Average optimality gap (%)**(backlog %, lost sales %, differentiation ratio)				
		GOP	**FCFS**	**DLP**	**SDP**	**SM_1**
All instances	1080	**0.00**(9.54, 17.72, 1.68)	**9.81**(0.00, 23.13, 1.00)	**15.98**(8.13, 24.57, 1.42)	**4.18**(12.77, 19.65, 1.57)	**6.07**(13.42, 21.04, 1.52)
Avg. backlogging (%) (Seg.1, Seg.2, Seg.3)		*(8.01, 9.66, 10.98)*	*(0.00, 0.00, 0.00)*	*(4.56, 7.73, 11.69)*	*(15.55, 13.82, 10.37)*	*(13.58, 13.90, 12.49)*
Avg. lost sales (%) (Seg.1, Seg.2, Seg.3)		*(1.20, 10.36, 41.31)*	*(22.94, 22.53, 22.95)*	*(11.49, 22.06, 37.67)*	*(4.56, 12.94, 39.13)*	*(8.81, 13.30, 40.05)*
CV = 1/3	270	**0.00**(9.07, 17.76, 1.85)	**9.27**(0.00, 22.26, 0.99)	**11.81** (10.20, 22.64, 1.61)	**3.24**(15.36, 19.14, 1.77)	**5.01**(16.41, 19.74, 1.69)
CV = 5/6	270	**0.00**(9.38, 17.67, 1.75)	**9.36**(0.00, 22.68, 1.01)	**14.42**(9.01, 23.65, 1.47)	**3.64**(13.71, 19.65, 1.58)	**5.76**(14.38, 20.55, 1.58)
CV = 4/3	270	**0.00**(8.93, 17.21, 1.63)	**9.76**(0.00, 22.40, 1.01)	**17.26**(6.88, 24.32, 1.36)	**4.63**(10.62, 18.96, 1.39)	**6.31**(11.80, 20.59, 1.44)
CV = 11/6	270	**0.00**(10.79, 18.22, 1.53)	**10.94**(0.00, 25.16, 0.99)	**21.04**(6.42, 27.67, 1.27)	**5.34**(11.40, 20.85, 1.30)	**7.36**(11.08, 23.28, 1.40)
r = (100,90,80)	360	**0.00**(9.08, 18.25, 1.61)	**6.50** (0.00, 23.88, 1.01)	**14.70**(8.21, 25.31, 1.42)	**3.12**(13.39, 20.07, 1.33)	**5.26**(16.45, 20.66, 1.17)
r = (100,80,60)	360	**0.00**(9.49, 17.29, 1.67)	**9.65**(0.00, 22.68, 0.99)	**15.78**(7.96, 24.16, 1.40)	**3.97**(13.18, 18.90, 1.49)	**5.86**(13.44, 20.36, 1.54)
r = (100,70,40)	360	**0.00**(10.05, 17.60, 1.78)	**13.99**(0.00, 22.82, 1.00)	**17.76**(8.21, 24.24, 1.44)	**5.68**(11.76, 19.98, 1.69)	**7.29**(10.37, 22.10, 2.02)
sr = 1%	360	**0.00**(10.27, 2.36, 1.06)	**7.76**(0.00, 8.86, 0.99)	**12.86**(6.96, 9.90, 1.08)	**2.75**(9.07, 5.00, 1.08)	**5.05**(9.45, 6,52, 1.12)
sr = 24%	360	**0.00**(11.40, 16.18, 1.68)	**9.76**(0.00, 22.55, 1.00)	**17.80**(11.27, 24.14, 1.37)	**4.64**(15.72, 18.70, 1.49)	**6.44**(15.82, 20.22, 1.54)
sr = 40%	360	**0.00**(6.96, 34.62, 4.25)	**11.36**(0.00, 37.98, 1.02)	**16.63**(6.15, 39.68, 2.32)	**4.80**(13.52, 35.25, 2.60)	**6.49**(15.00, 36.38, 2.55)

Table 10 Simulation results (continued)

Test bed subset	**N**	**Average optimality gap (%)**(backlog %, lost sales %, differentiation ratio)			
		SM_2	**DLP-BPC**	**RLP-BPC**	**DBPC**
All instances	1080	**5.46**(13.83, 20.47, 1.45)	**6.80**(16.76, 20.04, 1.42)	**6.90**(16.51, 20.20, 1.45)	**6.44**(11.64, 20.95, 1.56)
Avg. backlogging (%) (Cl.1, Cl.2, Cl.3)		*(16.70, 14.57, 10.09)*	*(18.38, 17.93, 13.68)*	*(18.14, 17.62, 13.52)*	*(13.32, 12.07, 9.25)*
Avg. lost sales (%) (Cl.1, Cl.2, Cl.3)		*(9.47, 13.17, 37.38)*	*(9.61, 13.35, 36.57)*	*(9.11, 13.61, 37.30)*	*(6.88, 14.82, 40.50)*
CV = 1/3	270	**4.20**(16.90, 19.11, 1.63)	**5.16**(19.77, 18.68, 1.62)	**4.75**(17.93, 19.11, 1.70)	**5.12**(10.98, 20.43, 1.82)
CV = 5/6	270	**5.09**(14.97, 19.94, 1.51)	**6.56**(18.09, 19.59, 1.47)	**6.44**(17.01, 19.95, 1.53)	**5.50**(9.75, 20.75, 1.61)
CV = 4/3	270	**5.75**(12.03, 20.07, 1.38)	**6.99**(14.67, 19.65, 1.36)	**7.50**(15.40, 19.72, 1.36)	**6.82**(12.09, 20.26, 1.51)
CV = 11/6	270	**6.97**(11.41, 22.76, 1.32)	**8.72**(14.49, 22.23, 1.27)	**9.19**(15.72, 22.03, 1.25)	**8.56**(13.72, 22.35, 1.36)
r = (100,90,80)	360	**5.26**(16.78, 20.58, 1.15)	**5.52**(16.56, 20.58, 1.26)	**5.64**(17.00, 20.61, 1.28)	**5.17**(11.71, 21.34, 1.43)
r = (100,80,60)	360	**5.37**(14.16, 19.84, 1.47)	**6.71**(16.34, 19.70, 1.46)	**6.88**(16.24, 19.83, 1.47)	**6.47**(12.15, 20.41, 1.56)
r = (100,70,40)	360	**5.79**(10.55, 20.99, 1.87)	**8.45**(17.38, 19.83, 1.58)	**8.44**(16.31, 20.17, 1.62)	**7.95**(11.05, 21.09, 1.73)
sr = 1%	360	**3.85**(9.07, 5.82, 1.09)	**4.87**(13.47, 4.70, 1.02)	**5.07**(12.83, 5.14, 1.05)	**4.95**(9.26, 6.04, 1.09)
sr = 24%	360	**5.80**(16.48, 19.56, 1.46)	**8.63**(22.40, 19.05, 1.35)	**8.09**(20.92, 19.24, 1.40)	**7.25**(13.82, 20.23, 1.57)
sr = 40%	360	**6.33**(15.93, 36.02, 2.39)	**6.56**(14.40, 36.36, 2.83)	**7.16**(15.80, 36.22, 2.75)	**6.81**(11.83, 36.57, 2.97)

From the first row of Table 10, we see that the SDP model, although theoretically no longer the optimal ex-ante policy, still performs best with an average optimality gap of 4.18%, followed by SM_2 and SM_1 with an average optimality gap of 5.46% and 6.07% respectively. The three bid-price control models show very close performance and the dynamic version performs slightly better than the two static versions. The DLP model performs much worse than the others with an optimality gap of 15.98%; it is much worse than even the FCFS policy. In summary, the main observations regarding the overall performance are: (1) the SDP model still performs best using a rolling planning horizon; (2) all the heuristics proposed perform close to each other, with SM_2 standing out a little.

Table 11 compares the optimality gaps of all the models with a finite planning horizon and a rolling planning horizon. For the three bid-price control models, we use the data of the resolved version (with a re-planning frequency of four periods) from the finite planning case.

Table 11 Comparison of optimality gap (%)

	FCFS	DLP	SDP	SM_1	SM_2	DLP-BPC	RLP-BPC	DBPC
Finite planning horizon	7.55	8.84	3.96	5.45	4.57	6.64	6.65	6.33
Rolling planning horizon	9.81	15.98	4.18	6.07	5.46	6.80	6.90	6.44

Compared to the finite horizon case, first, we find that the FCFS model's optimality gap increases from 7.55% in the finite horizon case to 9.81% in the rolling horizon case. As FCFS executes the same policy for both the finite and rolling planning horizons, it should generate the same absolute performance. Its increased optimality gap in the rolling horizon case then indicates that the GOP model works better with a rolling planning horizon because it makes its decisions based on full information for the total planning horizon.

From Table 11 we see that the DLP model performs much worse with a rolling planning horizon, the optimality gap being nearly double that of the finite horizon case. The potential explanation is that as the DLP model needs frequent re-planning to adjust its ATP allocation according to real consumption, a relatively low re-planning frequency of seven periods results in poor performance. To test this hypothesis, an additional simulation is run for the DLP model with a re-planning frequency of one period; this results in a drop in the optimality gap to 8.65%, which is quite close to the finite horizon case. However, the average run time increases from 2.56 seconds (see Table 9) to 7.37 seconds.

For the remaining models, the average performance with a rolling planning horizon is only a little worse than with the finite horizon. This might be due in part to the fact that the GOP itself is smarter now, which increases the optimality gaps. Overall, the comparison shows that the DLP is very sensitive to re-planning frequency, whereas for the other models it seems that the impact of re-planning frequency is limited.

I turn now to the lost sales percentage and backlogging behaviour of the models under different planning horizons. Comparing Table 10, Table 4 and Table 7, we can see that with a rolling planning horizon the safety margin models and the three bid-price control models have a much higher backlogging percentage (by approximately a factor of 3) compared to the finite horizon case. At the same time, the corresponding lost sales percentages decrease. Because of these two opposing effects, the overall performance of these models remains close to their performance in the finite horizon case. The DLP model shows the same pattern, i.e. the backlogging percentage increases while the lost sales rate decreases. However, compared to the other models, the increase in backlogging behaviour is not enough, which leads to a higher lost sales percentage and ultimately a bigger optimality gap.

Regarding the clear increase in backlogging behaviour in all the models, the possible reasons are twofold. First, in the finite horizon case, it is not possible to backlog in the second half of the planning horizon, after the second replenishment is delivered. However, in the rolling horizon case, backlogging is possible in almost any time period, i.e. there is a greater chance for the models to backlog. For example, the backlogging percentage of the GOP model increases from 3.42% in the finite horizon case to 9.54% in the rolling horizon case. Second, within each re-planning cycle with two replenishments, e.g. period 8 to 21, with available supply in periods 8 and 15, the model "sees" only half of the demand for the second supply cycle (demand from period 15 to 21) and ignores the other half (demand from period 22 to 28) as the next supply arrives only in period 29. Therefore, the model “thinks” that there is enough to backlog. If we extend the planning horizon length to enable the model to “see” the full demand for the second supply cycle, in principle this problem disappears. This hypothesis is tested in an additional simulation which extends the planning horizon length to 28 periods; the results are discussed later in this section.

From the second row of Table 10 we see that all models, except for the DLP model, backlog more of the higher classes than the lower classes. This is reasonable as it is only cost-effective to backlog the more profitable customers. However, the DLP model’s backlogging behaviour is strange as it backlogs even more for the lower classes. The possible explanation is that again, within each re-planning cycle with two replenishments, the model “sees” only half of the demand for the second supply cycle, resulting in some of the ATP quantities from the second supply not being allocated. According to the consumption rule of the DLP model, the unallocated ATP quantities can be consumed by orders from any customer class. As the on-hand supply allocated to Class 3 is very limited (especially when the shortage rate is high), the DLP model uses the unallocated ATP, which leads to a high backlogging percentage. Similar to the excessive backlogging issue, this problem might also disappear if we extend the planning horizon. Again, this hypothesis is tested using the additional simulation with an extended planning horizon length, the results of which are discussed later.

To summarize, compared to the finite horizon case, the rolling planning horizon seems to have no significant impact on the performance of the models. For the DLP model, the huge performance difference is mainly due to the different re-planning frequency.

In the above analysis, I argue that a longer planning horizon can have an impact on the performance of demand fulfilment models. In order to test this hypothesis, an additional simulation is conducted in which the planning horizon length is extended to 28 periods while keeping all other parameters unchanged for the computationally efficient models (namely, DLP, the safety margin models and the two static bid-price control models). In what follows, I provide a detailed analysis of the corresponding results summarized in Table 12.

Table 12 Simulation results (planning window = 28)

Test bed subset	N	Average optimality gap (%)(backlog %, lost sales %, differentiation ratio)				
		DLP_28	**SM_1_28**	**SM_2_28**	**DLP-BPC_28**	**RLP-BPC_28**
All instances	1080	**16.95**(3.36, 25.15, 1.45)	**9.22**(4.00, 24.65, 2.07)	**4.34**(8.36, 21.02, 1.46)	**5.98**(13.21, 19.75, 1.39)	**5.11**(11.52, 20.23, 1.44)
Avg. backlogging (%) (Seg.1, Seg.2, Seg.3)		*(1.77, 3.19, 4.86)*	*(4.91, 4.51, 2.35)*	*(12.49, 8.97, 3.55)*	*(16.23, 14.03, 9.22)*	*(14.48, 12.32, 7.61)*
Avg. lost sales (%) (Seg.1, Seg.2, Seg.3)		*(11.56, 22.20, 38.94)*	*(5.99, 12.54, 54.62)*	*(9.66, 14.20, 38.21)*	*(9.91, 13.44, 35.36)*	*(9.44, 13.54, 37.08)*
CV = 1/3	270	**11.34**(4.90, 22.53, 1.64)	**9.04** (3.62, 24.15, 2.47)	**2.53**(9.80, 19.54, 1.63)	**3.81**(15.13, 18.40, 1.53)	**3.29**(12.94, 18.84, 1.61)
CV = 5/6	270	**14.65**(3.76, 23.98, 1.49)	**9.38**(3.84, 24.50, 2.19)	**3.86**(8.52, 20.58, 1.50)	**5.12**(13.68, 19.36, 1.44)	**4.38**(11.54, 19.84, 1.51)
CV = 4/3	270	**18.75**(2.54, 25.15, 1.39)	**9.21**(3.77, 23.95, 1.93)	**4.88**(7.55, 20.57, 1.40)	**6.17**(11.24, 19.48, 1.34)	**5.55**(10.22, 19.85, 1.37)
CV = 11/6	270	**23.87**(2.23, 28.94, 1.28)	**9.25**(4.76, 26.01, 1.79)	**6.35**(7.60, 23.38, 1.34)	**9.13**(12.79, 21.77, 1.27)	**7.50**(11.39, 22.39, 1.28)
r = (100,90,80)	360	**15.32**(3.38, 25.87, 1.44)	**3.49**(6.56, 21.82, 1.36)	**3.41**(9.42, 21.20, 1.15)	**4.89**(12.55, 20.35, 1.21)	**3.95**(11.08, 20.75, 1.22)
r = (100,80,60)	360	**16.78**(3.41, 24.71, 1.43)	**9.49**(3.37, 24.20, 2.32)	**4.54**(9.01, 20.45, 1.48)	**5.88**(12.72, 19.50, 1.45)	**5.24**(11.38, 19.92, 1.49)
r = (100,70,40)	360	**19.10**(3.28, 24.87, 1.47)	**15.86**(2.07, 27.94, 3.21)	**5.25**(6.66, 21.40, 1.87)	**7.40**(14.37, 19.41, 1.56)	**6.37**(12.10, 20.02, 1.66)
sr = 1%	360	**12.70**(6.64, 9.21, 1.08)	**11.73**(2.72, 11.86, 1.37)	**4.11**(5.70, 6.80, 1.10)	**6.94**(15.85, 3.62, 1.01)	**4.55**(11.85, 4.94, 1.04)
sr = 24%	360	**19.27**(1.86, 25.87, 1.45)	**9.81**(4.40, 24.00, 2.18)	**4.46**(10.13, 20.08, 1.47)	**6.40**(15.37, 19.25, 1.33)	**5.82**(14.13, 19.42, 1.39)
sr = 40%	360	**17.96**(1.57, 40.37, 2.40)	**6.82**(4.87, 38.10, 4.15)	**4.40** (9.27, 36.18, 2.36)	**4.88**(8.41, 36.39, 2.69)	**4.89**(8.58, 36.34, 2.75)

From Table 12 we see that in extending the planning horizon to 28 periods the performance of SM_1 drops remarkably, while SM_2 even performs slightly better and thus the difference between them is greater. The performance of the two static bid-price control models is between the two safety margin models and the DLP model still performs worst with an optimality gap of 16.95%.

As expected, by extending the planning horizon, in general, the backlogging percentage of all models decreases sharply as now the models “see” the full demand for the second supply cycle. One might argue that although now the models “see” the full demand for the second supply cycle, the same problem still exists for the third supply cycle as when planning for periods 8 to 35, there are three available supplies in periods 8, 15 and 29. For the third supply arriving in period 29, the models also “see” only half of the demand for this supply cycle. However, as the re-planning frequency is 7 periods, it is not usually necessary to use the third supply for the first 7 periods, i.e. the allocation decisions regarding the third supply are not frozen (not implemented). Therefore, the problem regarding the third supply does not affect the backlogging behaviour of the models.

When the planning horizon length is 14 periods, the two safety margin models show nearly the same backlogging percentage. However, extending the planning horizon length to 28 periods, SM_1 backlogs much less than SM_2, which leads to higher lost sales. The potential explanation for their different behaviour is that when the planning horizon length is 14 periods, as the models “see” only half of the demand for the second supply, the “double-counting” effect of SM_1 is not severe. When the planning horizon length is 28 periods, SM_1 has not only a “double-counting” problem, but even a “triple-counting” problem. So, for example, when considering the orders that arrive before the second supply, compared to SM_2, SM_1 allocates more second supply to the future customers, which significantly limits the backlogging possibility for the current order. In general, due to the “triple-counting” effect, SM_1 over-protects the future higher classes, which can also be seen in the third row of Table 12, and therefore performs much worse than SM_2. The above analysis shows that SM_1 is sensitive to the choice of parameters of the rolling planning scheme, which determine how much future demand is seen or not seen by the models.

Based on the above analysis, we can draw a general conclusion concerning SM_1: the longer the planning horizon, the worse SM_1 performs. The reason for this is that due to the “double-counting” effect, the longer planning horizon induces SM_1 to reserve more supply for the higher classes until it reaches a point at which the holding cost prevents it reserving further. Thus, in order to make use of this method, the planning horizon should be kept short.

Regarding the strange backlogging behaviour of the DLP model, on the one hand we see that by extending the planning horizon to 28 periods, the model shows much less backlogging for the lower classes. On the other hand, the backlogging percentage is still greater for the lower classes than for the higher classes. This is because, although for the second supply cycle the model “sees” the full demand, for the third cycle it “sees” only half of the demand. As a deterministic model, the DLP model allocates over the whole planning window; for this window, when expected demand is less than the available supply, there may still be unallocated ATP quantities in the second supply, which leads to backlogging for the lower classes. Thus, based on the above analysis, we can see that similar to SM_1, the DLP model is also sensitive to the choice of parameters in the rolling planning scheme.

5.3.2.2 Sensitivity Analysis

The second part of Table 10 provides information on the impact of different design factors on the performance of each fulfilment model. In what follows, I discuss the impact of the coefficient of variation of the order size (CV), customer heterogeneity (r) and supply shortage rate (sr).

From Table 10 and Figures 10 to 12, we can see that the impact of the design factors with a rolling planning horizon is quite consistent with the impact in the finite horizon case. As the CV value increases, all strategies show an increasing trend in their average optimality gaps. Demand uncertainty has greatest impact on the DLP model: as the CV value increases, the optimality gap increases sharply. As the scale of customer heterogeneity increases, the performance of all strategies decreases and FCFS is most affected by increasing heterogeneity. For most of the models, the impact of the shortage rate is not monotonic. They perform worst for an intermediate shortage rate of 24%. The performance of SM_2 shows a decreasing pattern as the shortage increases.

From the overall picture, with the smallest optimality gap, the SDP model still performs best among all the other methods, but it is also computationally the most expensive. The DLP model performs worst. The performance of all the proposed models is close and SM_2 delivers the closest approximation to the SDP model.

Figure 10 Average optimality gap for different CV values

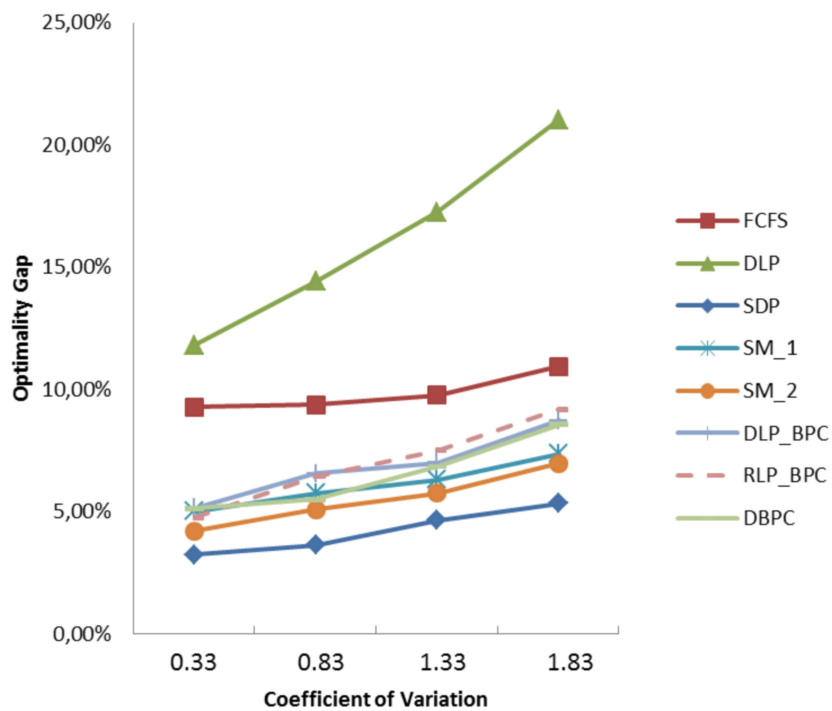


Figure 11 Average optimality gap for different customer heterogeneity

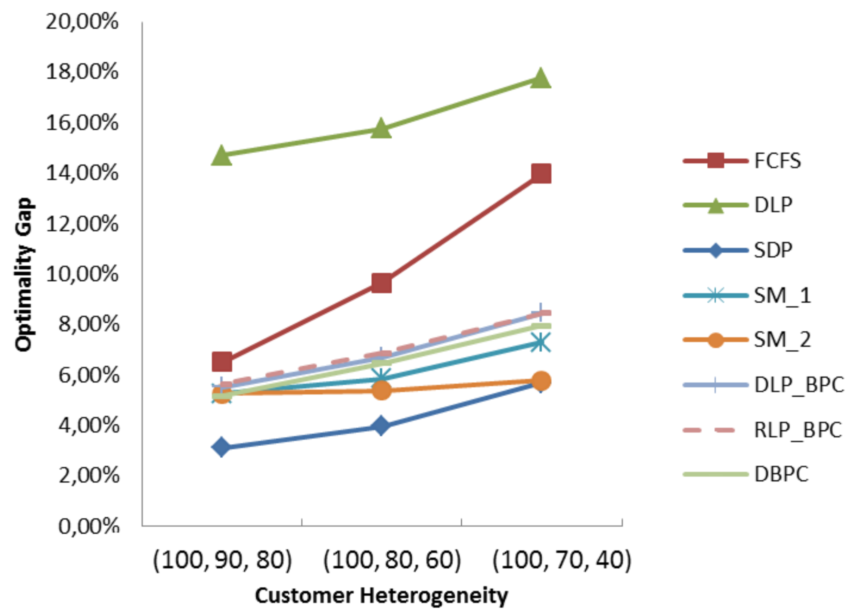
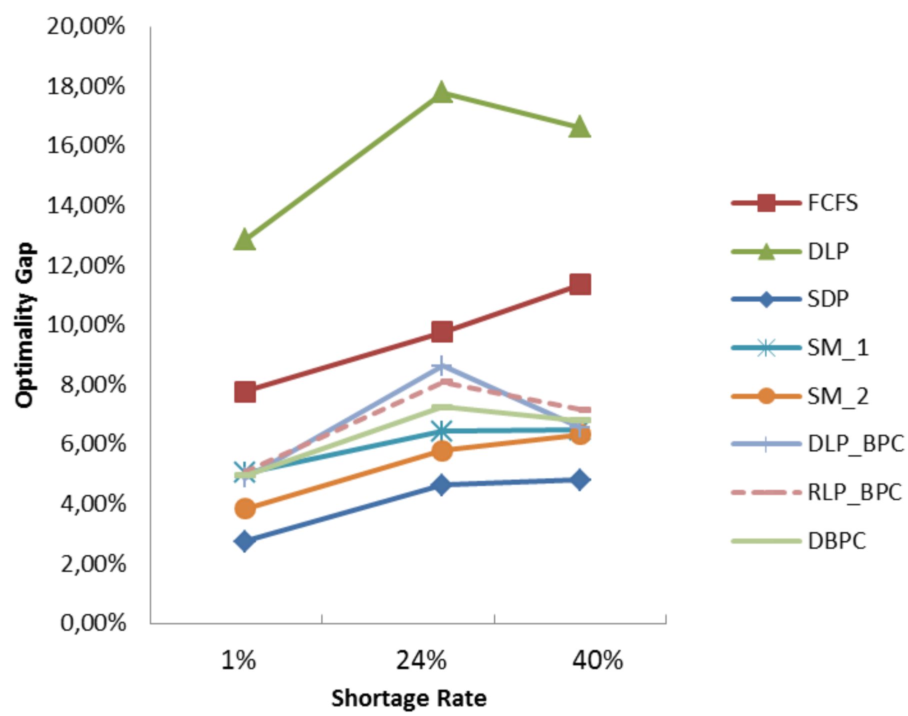


Figure 12 Average optimality gap for different supply scarcity



5.4 Summary

So far, rolling horizon planning has not gained much attention in the revenue management research community, largely because in the current major area of application, i.e. service industries, there tends to be a natural end to the planning horizon, e.g. the take-off time in the case of airlines. However, as explained in the introduction to this chapter, rolling horizon planning is common practice in manufacturing industries and therefore it is necessary to extend the revenue management models to adapt to rolling horizon planning.

In this chapter, each re-planning cycle has simply been treated as a finite planning horizon problem and the corresponding models have been implemented as if they related to a finite planning horizon. The numerical results show that with a rolling planning horizon, revenue management approaches still make sense as all of the proposed models significantly outperform the simple FCFS policy. Of all these, the SDP model is still the best-performing model with a rolling planning horizon, although theoretically it is no longer the optimal ex-ante policy. As expected, scalability is still the main problem with the SDP model. All the other demand fulfilment models considered are much more efficient than the SDP model, especially the DLP model, the safety margin models and the two static bid-price control models.

However, SM_1 is sensitive to the choice of parameters in the rolling planning scheme, determining how much future demand is seen or not seen by the models. The DLP model has the same problem and is also sensitive to the choice of re-planning frequency. For the other models, there is no indication that the choice of parameter has a significant impact on performance. Of these models, SM_2 provides the closest approximation to the SDP model and therefore has considerable potential for practical application.

In general, for manufacturing industries, rolling horizon planning resembles reality better than finite horizon planning and this study shows that the models proposed generate similar results using rolling horizon planning as for finite horizon planning. In other words, the conclusions drawn in the finite case are still valid in the rolling horizon case. Therefore, we can conclude that the models proposed can be used in rolling horizon planning.

Chapter VI

Conclusion

In this thesis, revenue management approaches have been applied to demand fulfilment problems in a make-to-stock manufacturing (MTS) system. The topic is motivated by the demand fulfilment task pertinent to the advanced planning systems (APS) used today, in which the decision maker has to decide how to allocate the limited available-to-promise (ATP) quantities to different customer classes to maximize profits.

In APS, the ATP quantities are derived from mid-term master planning and cannot be changed in the short term. This resembles the traditional revenue management problem, in which a fixed amount of perishable assets is sold to multiple customer classes to maximize revenue. However, the difference is that the ATP quantities are not perishable and can be replenished at certain times. Therefore, the objective here is no longer to maximize revenue, but the overall profit, taking into account the inventory holding cost and backlogging cost.

In Chapter 2, the exact problem setting is defined: an MTS manufacturer is facing stochastic demand from heterogeneous customers. To maximize the expected profit, the manufacturer has to decide whether to satisfy each arriving order from stock, backorder it at a penalty cost, or reject it in anticipation of more profitable future orders. The replenishments are exogenously determined and the manufacturer needs to take into account not only sales revenues, but also inventory holding costs and back-order penalties. A common mathematical model is then set up to study this demand fulfilment problem. Two existing models are revisited, namely a stochastic dynamic programming (SDP) model (Quante et al., 2009) and a deterministic linear programming (DLP) model (Meyr, 2009). The SDP model provides the optimal ex-ante policy for the demand fulfilment model, but due to its high computational effort, it is scarcely applicable in real-life practice. The DLP model, on the other hand, is efficient to solve. However, as it ignores demand uncertainty, the solution is usually suboptimal. To overcome the limitations of the two existing models, new approaches are developed in the following chapters of the thesis.

To incorporate demand uncertainty into the DLP model, I borrow the safety stock idea from inventory management. In Chapter 3, I develop two versions of a safety margin model, which reserve certain stock as a “safety margin” for more profitable customers. Following the two-level planning procedure of APS, in the allocation planning level, I use the idea of the expected marginal seat revenue (EMSR) heuristic to calculate safety margins. However, EMSR deals with only one single resource and assumes that low-revenue demand arrives before high-revenue demand. These assumptions are not valid in the MTS setting. Therefore, to make use of the EMSR, I consider the multiple ATP supplies separately and rank customers according to their arrival date and unit revenue. In the calculation, I ensure that only the future higher classes are protected. Finally, the safety margins obtained are used to calculate the corresponding booking limits, which are used in the order promising level. The difference between the two versions of the safety margin model is that when calculating safety margins using EMSR, SM_1 takes all future demand into consideration, whereas SM_2 only considers future demand that arrives before the next ATP supply. Due to the fact that the safety margin calculation is independent of the real consumption of the ATP quantities, it is not necessary for the safety margin models to repeat the allocation planning steps before each order processing. A numerical study shows that the safety margin models are computationally efficient and improve substantially on the performance of the pure DLP model. They even perform very close to level of the SDP model. The safety margin models contribute to linking the traditional inventory/supply chain management world to the emerging revenue management world. The main limitation of the safety margin models is that as the different supplies are considered separately in the allocation stage, there is “double counting” of the demand of the higher classes in SM_1, which makes the model over-protect the more valuable customers. In contrast, there may be insufficient protection for the higher classes in SM_2 as only a fraction of the demand is considered. Therefore, in future research it would be worth considering different approaches to calculate the safety margins.

To overcome the computational intractability of the SDP model, in Chapter 4 bid-price control is used to approximate it. The basic idea is to approximate dynamic programming (DP) using an efficient mathematical programming formulation, e.g. linear programming (LP), and solve the dual problem to obtain the shadow prices which are then considered bid prices (Bertsimas & Popescu, 2003). In the literature review section of Chapter 4, first the similarities and differences between the demand fulfilment problem and traditional network revenue management problems are compared. Building on the insights from this comparison, three bid-price control models are then developed. For the DLP-based bid-price control model and the RLP-based bid-price control model, the ideas are the same: to discard the primal solutions of the original model (Meyr, 2009; Quante, 2008) and calculate the corresponding bid prices based on the associated dual prices. For the dynamic bid-price control model, following Adelman (2007), an affine functional approximation is made to the value function of the SDP model. What makes the dynamic model different from Adelman's (2007) model is that unlike in the airline case, it is necessary to decide not only whether to satisfy a given order or not, but also which supply to use and how much of each supply to use. As reflected in the modelling, the decision variable is no longer a binary variable indicating whether or not to accept a certain order, but an integer vector denoting different ATP quantities used to satisfy the incoming order. Solving the dual problem of the LP formulation of the approximated DP model using column generation, we obtain a time trajectory of bid prices all at once. Following the same numerical study framework as in Chapter 3, the performance of the three proposed bid-price control models is compared. As the best-performing bid-price control model, the dynamic model provides a close approximation to the SDP model with much lower computational effort. Without frequent resolving, it performs substantially better than the two static models; with resolving, all three models generate similar performance. One limitation of the proposed dynamic model is that it captures only the temporal dynamics of demand but ignores the impact of remaining capacity. Therefore, for future research, it is worth considering dynamic models that generate both time-dependent and capacity-dependent bid prices.

Finally, in Chapter 5 the analysis is extended to rolling horizon planning as this is common practice in manufacturing industries. As this study is the first step towards applying revenue management in rolling horizon planning, each re-planning cycle of the rolling horizon system is simply treated as a finite planning horizon problem and applied to the various demand fulfilment models. Based on literature that studies the impact of different policy decision variables which define the implementation strategies for rolling horizon planning, the rolling planning schemes are fixed and a series of numerical studies are conducted to analyse the performance of the demand fulfilment models with a rolling planning horizon. The results show that using a rolling planning horizon, the revenue management idea still makes sense as all the models, except the DLP model, generate much better performance than the FCFS policy. The SDP model still provides the best solution with a rolling planning horizon, although theoretically it is no longer the optimal ex-ante policy. However, as in the finite case, it is too expensive computationally to apply SDP in practice. The choice of the policy decision variables (planning horizon length, frozen interval and re-planning frequency) does have an impact on the performance of models. Among the demand fulfilment models, we find that the DLP model and SM_1 are the most sensitive. Of all the efficient methods, SM_2 provides the closest performance to the SDP model and appears to be robust with respect to the parameter choice of the rolling horizon system. Therefore, we can conclude that it has high potential for practical application. The main limitations of the numerical study are that the re-planning frequency is not changed for all models and only two levels of planning horizon length are considered. In future research, it would be worth conducting a more comprehensive numerical study, for example based on a full factorial design, to analyse the detailed impact of the policy decision variables on the performance of different models.

Based on revenue management ideas, this thesis examines the development of new models that overcome the limitations of two existing optimization models for demand fulfilment in an MTS manufacturing system. As the common mathematical model is based on the planning framework of APS, the models proposed can easily be adapted to current APS practice. The main limitation of this thesis is that in the common mathematical model, it is assumed that the order due date is equal to the order arrival date. For future research, it would be interesting to extend the proposed models to include different customer due dates and see the impact on the overall performance of the models.

Another interesting future research direction would be to apply revenue management approaches to an assemble-to-order (ATO) manufacturing system. Nowadays, as mass customization is becoming increasingly popular, many companies are shifting from an MTS system to an ATO system. Therefore, it would be worth extending these models to an ATO system. Unlike in an MTS system, in an ATO system, one has to allocate both components and assembly capacity. Whereas components are storable and can be replenished, assembly capacity is perishable. This hybrid feature makes revenue management application in an ATO system even more challenging.

References

Adelman, D. (2007). Dynamic bid prices in revenue management. *Operations Research, 55*, 647-661.

Adelman, D., & Mersereau, A. J. (2008). Relaxations of weakly coupled stochastic dynamic programs. *Operations Research, 56*(3), 712-727.

Baker, K. R. (1977). An experimental study of the effectiveness of rolling schedules in production planning. *Decision Sciences, 8*(1), 19-27.

Ball, M., Chen, C.-Y., & Zhao, Z.-Y. (2004). Available to promise. In D. Simchi-Levi, S. Wu & Z.-J. Shen (Eds.), *Handbook of quantitative supply chain analysis* (Vol. 74, pp. 447-483). Springer US.

Bardhan, A., Dawande, M., Gavirneni , S., Mu, Y., & Sethi, S. (2013). Forecast and rolling horizons under demand substitution and production changeovers: Analysis and insights. *IIE Transactions, 45*(3), 323-340.

Barut, M., & Sridharan, V. (2004). Design and evaluation of a dynamic capacity apportionment procedure. *European Journal of Operational Research, 155*(1), 112-133.

Barut, M., & Sridharan, V. (2005). Revenue management in order-driven production systems. *Decision Sciences, 36*(2), 287-316.

Bearden, J. N., Murphy, R. O., & Rapoport, A. (2008). Decision biases in revenue management: Some behavioral evidence. *Manufacturing & Service Operations Management, 10*(4), 625-636.

Belobaba, P. (2001). Revenue and competitive impacts of OD control: Summary of PODS results. In *First Annual INFORMS Revenue Management Section Meeting, New York, NY*.

Belobaba, P. P. (1987). Airline yeld management. An overview of seat inventory control. *Transportation Science, 21*(2), 63.

Belobaba, P. P. (1989). OR practice – Application of a probabilistic decision model to airline seat inventory control. *Operations Research, 37*(2), 183-197.

Belobaba, P., & Lee, S. (2000). PODS update: Large network OD control results. In *Proc. 2000 AGIFORS Reservations and Yield Management Study Group Symposium. New York, NY.*

Bendoly, E. (2011). Linking task conditions to physiology and judgment errors in RM systems. *Production and Operations Management, 20*(6), 860-876.

Bertsimas, D., & Popescu, I. (2003). Revenue management in a dynamic network environment. *Transportation Science, 37*, 257-277.

Blackburn, J. D., & Millen, R. A. (1982). The impact of a rolling schedule in a multi-level MRP system. *Journal of Operations Management, 2*(2), 125-135.

Bookbinder, J. H., & H'ng, B.-T. (1986). Production lot sizing for deterministic rolling schedules. *Journal of Operations Management, 6*(3), 349-362.

Bront, J., Méndez-Díaz, I., & Vulcano, G. (2009). A column generation algorithm for choice-based network revenue management. *Operations Research, 57*, 769-784.

Brown, D. B., Smith, J. E., & Sun, P. (2010). Information relaxations and duality in stochastic dynamic programs. *Operations Research, 58*(4-Part-1), 785-801.

Carlson, R. C., Beckman, S. L., & Kropp, D. H. (1982). The effectiveness of extending the horizon in rolling production scheduling. *Decision Sciences, 13*(1), 129-146.

Carlson, R. C., & Yano, C. A. (1986). Safety stocks in MRP – Systems with emergency setups for components. *Management Science, 32*(4), 403-412.

Chung, C.-H., & Krajewski, L. J. (1984). Planning horizons for master production scheduling. *Journal of Operations Management , 4*(4), 389-406.

De Boer, S. V., Freling, R., & Piersma, N. (2002). Mathematical programming for network revenue management revisited. *European Journal of Operational Research, 137*, 72-92.

Defregger, F., & Kuhn, H. (2004). Revenue management in manufacturing. In D. Ahr, R. Fahrion, M. Oswald & G. Reinelt (Eds.), *Operations research proceedings 2003* (Vol. 2003, pp. 17-22). Berlin, Heidelberg: Springer.

Defregger, F., & Kuhn, H. (2007). Revenue management for a make-to-order company with limited inventory capacity. *OR Spectrum, 29*(1), 137-156.

Erdelyi, A., & Topaloglu, H. (2010). A dynamic programming decomposition method for making overbooking decisions over an airline network. *INFORMS J. on Computing, 22*, 443-456.

Filho, M. G., & Fernandes, F. F. (2009). A practical method to reduce nervousness and improve performance in MRP systems: Proposal, implementation and results. *Production and Inventory Management Journal, 45*(2), 65-79.

Fleischmann, B., & Meyr, H. (2003). Planning hierarchy, modeling and advanced planning systems. *Handbooks in Operations Research and Management Science, 11*, 455-523.

Gallego, G., Iyengar, G., Phillips, R., & Dubey, A. (2004). Managing flexible products on a network. *CORC Technical Report TR-2004-01, Columbia University Department of Industrial Engineering and Operations Research, New York*

Gallego, G., & Phillips, R. (2004). Revenue management of flexible products. *Manufacturing & Service Operations Management, 6*(4), 321-337.

Harris deB., F. H., & Pinder, J. P. (1995). A revenue management approach to demand management and order booking in assemble-to-order manufacturing. *Journal of Operations Management , 13*(4), 299-309.

Jung, H. (2010). An available-to-promise model considering customer priority and variance of penalty costs. *The International Journal of Advanced Manufacturing Technology, 49*, 369-377.

Kilger, C., & Meyr, H. (2008). Demand fulfilment and ATP. In H. Stadtler & C. Kilger (Eds.), *Supply chain management and advanced planning* (pp. 181-198). Berlin, Heidelberg: Springer.

Klein, R. (2007). Network capacity control using self-adjusting bid-prices. *OR Spectrum, 29*(1), 39-60.

Klein, R., & Steinhardt, C. (2008). *Revenue management: Grundlagen und Mathematische Methoden.*Berlin: Springer.

Kunnumkal, S., & Topaloglu, H. (2010). Computing time-dependent bid prices in network revenue management problems. *Transportation Science, 44*, 38-62.

Lian, Z., Liu, L., & Zhu, S. X. (2010). Rolling-horizon replenishment: Policies and performance analysis. *Naval Research Logistics (NRL), 57*(6), 489-502.

Lin, N.-P., Krajewski, L., Leong, G., & Benton, W. (1994). The effects of environmental factors on the design of master production scheduling systems. *Journal of Operations Management, 11*(4), 367-384.

Maxwell, M. S., Restrepo, M., Henderson, S. G., & Topaloglu, H. (2010). Approximate dynamic programming for ambulance redeployment. *INFORMS Journal on Computing, 22*(2), 266-281.

Meyr, H. (2009). Customer segmentation, allocation planning and order promising in make-to-stock production. *OR Spectrum, 31*, 229-256.

Meyr, H., Wagner, M., & Rohde, J. (2008). Structure of advanced planning systems. In H. Stadtler & C. Kilger (Eds.), *Supply chain management and advanced planning* (pp. 109-115). Berling, Heidelberg: Springer.

Narayanan, A., & Robinson, P. (2010). Evaluation of joint replenishment lot-sizing procedures in rolling horizon planning systems. *International Journal of Production Economics , 127*(1), 85-94.

Nathan, J., & Venkataraman, R. (1998). Determination of master production schedule replanning frequency for various forecast window intervals. *International Journal of Operations & Production Management, 18*(8), 767-777.

Perakis, G., & Roels, G. (2010). Robust controls for network revenue management. *Manufacturing & Service Operations Management, 12*(1), 56-76.

Petrick, A., Gönsch, J., Steinhardt, C., & Klein, R. (2010). Dynamic control mechanisms for revenue management with flexible products. *Computers & Operations Research, 37*(11), 2027-2039.

Petrick, A., Steinhardt, C., Gönsch, J., & Klein, R. (2012). Using flexible products to cope with demand uncertainty in revenue management. *OR Spectrum, 34*(1), 215-242.

Pibernik, R. (2005). Advanced available-to-promise: Classification, selected methods and requirements for operations and inventory management. *International Journal of Production Economics, 93*, 239-252.

Pibernik, R. (2006). Managing stock-outs effectively with order fulfilment systems. *Journal of Manufacturing Technology Management, 17*(6), 721-736.

Pibernik, R., & Yadav, P. (2009). Inventory reservation and real-time order promising in a make-to-stock system. *OR Spectrum, 31*, 281-307.

Quante, R. (2008). *Management of stochastic demand in make-to-stock manufacturing.* Ph.D. Thesis, Institute for Production Management, Vienna University of Economics and Business Administration, Vienna.

Quante, R., Fleischmann, M., & Meyr, H. (2009). A stochastic dynamic programming approach to revenue management in a make-to-stock production system. *Working Paper*.

Sahin, F., Narayanan, A., & Robinson, E. P. (2013). Rolling horizon planning in supply chains: Review, implications and directions for future research. *International Journal of Production Research, 51*(18), 5413-5436.

Secomandi, N. (2008). An analysis of the control-algorithm re-solving issue in inventory and revenue management. *Manufacturing & Service Operations Management, 10*(3), 468-483.

Sethi, S., & Sorger, G. (1991). A theory of rolling horizon decision making. *Annals of Operations Research, 29*(1), 387-415.

Shen, Z.-J. M., & Su, X. (2007). Customer behavior modeling in revenue management and auctions: A review and new research opportunities. *Production and Operations Management, 16*(6), 713-728.

Simpson, N. (1999). Multiple level production planning in rolling horizon assembly environments . *European Journal of Operational Research , 114*(1), 15-28.

Spengler, T., & Rehkopf, S. (2005). Revenue management konzepte zur entscheidungsunterstützung bei der annahme von kundenaufträgen. *Zeitschrift für Planung & Unternehmenssteuerung, 16*(2), 123-146.

Spengler, T., Rehkopf, S., & Volling, T. (2007). Revenue management in make-to-order manufacturing – An application to the iron and steel industry. *OR Spectrum, 29*, 157-171.

Spengler, T., Volling, T., Wittek, K., & Akyol, D. (2009). Dynamic bid-price policies for make-to-order revenue management. In B. Fleischmann, K.-H. Borgwardt, R. Klein & A. Tuma (Eds.), *Operations research proceedings 2008* (Vol. 2008, pp. 103-108). Berlin, Heidelberg: Springer.

Sridharan, V., & Berry, W. L. (1990a). Master production scheduling make-to-stock products: A framework for analysis. *International Journal of Production Research, 28*(3), 541-558.

Sridharan, V., & Berry, W. L. (1990b). Freezing the master production schedule under demand uncertainty. *Decision Sciences, 21*(1), 97-120.

Sridharan, V., Berry, W. L., & Udayabhanu, V. (1987). Freezing the master production schedule under rolling planning horizons. *Management Science, 33*(9), 1137-1149.

Sridharan, S. V., Berry, W. L., & Udayabhanu, V. (1988). Measuring master production schedule stability under rolling planning horizons. *Decision Sciences, 19*(1), 147-166.

Sridharan, S., & LaForge, R. (1990). An analysis of alternative policies to achieve schedule stability. *Journal of Manufacturing and Operations Management, 3*(1), 53-73.

Sridharan, V., & LaForge, R. L. (1994). Freezing the master production schedule: implications for fill rate. *Decision Sciences, 25*(3), 461-469.

Talluri, K., & van Ryzin, G. (1998). An analysis of bid-price controls for network revenue management. *Management Science, 44*(11), pp. 1577-1593.

Talluri, K., & van Ryzin, G. (1999). A randomized linear programming method for computing network bid prices. *Transportation Science, 33*(2), 207-216.

Talluri, K., & van Ryzin, G. (2004a). *The theory and practice of revenue management.* Dordrecht: Kluwer Academic Publishers.

Talluri, K., & van Ryzin, G. (2004b). Revenue management under a general discrete choice model of consumer behavior. *Management Science, 50*(1), pp. 15-33.

Topaloglu, H. (2009). Using Lagrangian relaxation to compute capacity-dependent bid prices in network revenue management. *Operations Research, 57*, 637-649.

van Ryzin, G., & Vulcano, G. (2008). Simulation-based optimization of virtual nesting controls for network revenue management. *Operations Research, 56*(4), 865-880.

van Slyke, R., & Young, Y. (2000). Finite horizon stochastic knapsacks with applications to yield management. *Operations Research, 48*(1), pp. 155-172.

Venkataraman, R., & D'Itri, M. P. (2001). Rolling horizon master production schedule performance: A policy analysis. *Production Planning & Control, 12*(7), 669-679.

Vulcano, G., van Ryzin, G., & Chaar, W. (2010). OM practice-choice-based revenue management: An empirical study of estimation and optimization. *Manufacturing & Service Operations Management, 12*, 371-392.

Williamson, E. L. (1992). *Airline network seat control.* Ph.D. thesis, MIT, Cambridge, Massachusetts.

Xiao, Y., Chen, J., & Lee, C.-Y. (2010). Single-period two-product assemble-to-order systems with a common component and uncertain demand patterns. *Production and Operations Management, 19*(2), 216-232.

Xie, J., Zhao, X., & Lee, T. (2003). Freezing the master production schedule under single resource constraint and demand uncertainty. *International Journal of Production Economics, 83*(1), 65-84.

Yano, C. A., & Carlson, R. C. (1985). An analysis of scheduling policies in multiechelon production systems. *IIE Transactions, 17*(4), 370-377.

Yano, C. A., & Carlson, R. C. (1987). Interaction between frequency of rescheduling and the role of safety stock in material requirements planning systems. *International Journal of Production Research, 25*(2), 221-232.

Zhang, D. (2010). An improved dynamic programming decomposition approach for network revenue management. *Manufacturing & Service Operations Management, 13*(1), 35-52.

Zhang, D., & Adelman, D. (2009). An approximate dynamic programming approach to network revenue management with customer choice. *Transportation Science, 43*(3), 381-394.

Zhao, X., & Lam, K. (1997). Lot-sizing rules and freezing the master production schedule in material requirements planning systems. *International Journal of Production Economics, 53*(3), 281-305.

Zhao, X., & Lee, T. (1993). Freezing the master production schedule for material requirements planning systems under demand uncertainty. *Journal of Operations Management, 11*(2), 185-205.

Zhao, X., & Lee, T. S. (1996). Freezing the master production schedule in multilevel material requirements planning systems under deterministic demand. *Production Planning & Control, 7*(2), 144-161.

Zhao, X., Goodale, J. C., & Lee, T. S. (1995). Lot-sizing rules and freezing the master production schedule in material requirements planning systems under demand uncertainty. *International Journal of Production Research, 33*(8), 2241-2276.

Curriculum Vitae

Yao Yang

10.2010 – 11.2014	University of Mannheim, Germany Doctoral Candidate, Research Assistant at the Chair of Logistics and Supply Chain Management
09.2008 – 09 .2009	University of Mannheim, Germany Doctoral Candidate in Operations & Information Systems at the Center for Doctoral Studies in Business
03.2006 – 08.2007	RWTH Aachen, Germany Master in Production System Engineering
09.2005 – 07.2008	Tsinghua University, China Master in Management Science & Engineering
09.2001 – 07.2005	Tsinghua University, China Bachelor in Management Science & Engineering