UNIVERSITY OF MANNHEIM

Bid price-based revenue management approaches in manufacturing industries

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

vorgelegt von

Dipl.-Math. Hendrik Guhlich
Mannheim

Dekan:	*Prof. Dr. Dieter Truxius*
Referent:	*Prof. Dr. Raik Stolletz*
Korreferent:	*Prof. Dr. Moritz Fleischmann*

Tag der mündlichen Prüfung: *29. April 2016*

Summary

Whenever demand exceeds capacity, the available resources must be allocated to the incoming demand. A simple first-come-first-served allocation can lead to rather poor results if the demand is heterogeneous, i.e., if orders differ in their strategic importance for a company, their willingness to pay, their lead time requirements, etc. In addition, the decisions must be made under uncertainty because demand is often stochastic. If too many low value orders are rejected, capacity might remain unused. On the other hand, if too many low value orders are accepted, there might be not enough capacity to accept all high value orders. Apart from the order acceptance decision, it might also be preferable to delay the order release of already accepted orders, even if capacity is available, to reserve capacity for future high value orders.

This thesis considers the demand management decisions in different order-driven production environments. Orders are accepted or rejected immediately upon arrival, and at the beginning of each planning period, the set of orders to be released is selected. As exact solution methods are computationally intractable, bid price-based revenue management approaches are applied to maximize the contribution margin gained by accepted orders minus holding and backlog costs.

The first essay considers a deterministic single-stage assembly process, where both capacity and intermediate materials are scarce resources. In this essay, for each accepted order a due date is quoted and the accepted orders are scheduled such that the quoted due dates are kept. The second essay considers a deterministic multi-stage make-to-order production system. In this study, intermediate materials are assumed to be unlimited available. The third essay considers a multi-stage make-to-order production system with stochastic influences. To take into account that in these production systems, lead times depend non-linearly on the load of the system, clearing functions are used to model the production system.

Contents

List of Figures

List of Tables

1 Introduction

Whenever demand exceeds capacity, the available resources must be allocated to the customers. This allocation is especially important if these customers are heterogeneous, i.e., they differ in their importance for the company, willingness to pay, lead time requirements, etc. What makes this problem difficult to solve is that demand is typically uncertain. The most famous example is the sale of flight tickets:

The number of seats on a flight is fixed and demand can be roughly divided into business and leisure travelers. Business travelers typically book their flight on short notice and are less price-sensitive than leisure travelers. As demand cannot be forecasted exactly, the question arises, how many low price tickets to sell to early arriving customers. If too many low price tickets are sold, later arriving high value customers must be rejected. If too many requests for low price tickets are rejected, available seats remain unsold although sufficient demand was available.

Revenue management approaches have been developed to address this problem (see Talluri and Van Ryzin, 2004, for an overview) and, in recent years, this concept has also been adapted to production industries (see e.g. Spengler et al., 2007; Quante et al., 2009). The focus in this thesis is on applications in different order-driven production environments.

One possible field of application can be found in the semiconductor industry. Here, the production process is divided into two stages: front-end and back-end production. In the front-end production, wafers are fabricated, tested, and stored in the so called die bank. In back-end facilities, the wafers are taken from the die bank, placed into packages, and finally tested (Brown et al., 2000; Mönch et al., 2013). For this thesis, the back-end facilities are of interest because here, the demand management decisions are made. Demand is heterogeneous and characterized by demand peaks. Capital equipment costs are high and thus, it is impossible to adjust capacity levels in the short-term. Therefore, demand exceeds capacity in the short-term. The production characteristics vary for different products. For well established products, which have been produced for a long time, processing times can be assumed to be almost deterministic, while for new developed products, they may be highly stochastic. In addition, for some products, all required intermediate materials are available or their lead time is negligible, while for other products, they are scarce resources.

Another possible field of application is in the steel industry, where different products are produced to order in a multi-stage no-wait flow shop (Spengler et al., 2007, 2008). Capacity cannot be extended in the short-term because this is very costly and time-consuming. Orders can be rejected because there are many customers without a long-term contract whose overall economic impact is relatively low (Hintsches et al., 2010).

This thesis considers the order acceptance and order release decisions of a manufacturer facing stochastic, heterogeneous demand. In each of the three articles, a different production environment is analyzed. As exact solution methods are computationally intractable, bid price-based revenue management approaches are developed to maximize the contribution margin gained from accepted orders minus backlog and holding costs.

The first article (joined work with Moritz Fleischmann and Raik Stolletz)[1] considers an assemble-to-order production system, where both intermediate materials and production capacity are scarce resources. Stochastic, heterogeneous demand arrives over time. Upon arrival, each order must be either accepted or rejected. For each accepted order, a due date is quoted, which then must be kept. The actual assembly date is still subject to change. At the beginning of each planning period, the set of orders for which assembly is started in the current planning period is selected. Orders must be fully accepted but partial assembly starts are possible. The objective is to maximize the contribution margin gained from accepted orders minus backlog costs, which depend on the quoted due date, and holding costs, which depend on the actual assembly date.

This problem is formulated as a stochastic dynamic program to show differences to other approaches from the literature, where scarce intermediate materials are not considered and different cost structures are assumed. Due to the high dimensionality of the resulting state space for problems of realistic size, the stochastic dynamic program is computationally intractable to solve. Therefore, a bid price-based revenue management approach is developed. Bid prices are derived from a linear program describing the production process. They are used both in the due date quoting decision and in the scheduling decision. For the former decision, the bid prices are aggregated to opportunity costs of accepting an order with a given due date and for the latter decision, they are aggregated to opportunity costs of starting assembly of an order in a given planning period.

[1]Guhlich, H., Fleischmann, M., and Stolletz, R. (2015). Revenue management approach to due date quoting and scheduling in an assemble-to-order production system. OR Spectrum, 37(4):951-982.

A numerical study shows that the proposed approach performs well by comparing it with benchmark algorithms and an ex-post optimal solution, which has knowledge about all incoming demand. In addition, it is shown that considering production capacity is more important than considering intermediate materials to get good results in the tested cases.

The second article (joined work with Moritz Fleischmann, Lars Mönch, and Raik Stolletz)[2] considers a multi-stage no-wait production system. The make-to-order production principle is used and intermediate materials are unlimited available in this case. Demand is stochastic and each incoming order must be fully accepted or rejected immediately on arrival. In addition, order release decisions are performed periodically. These decisions are made online. That is, decisions from former periods can be revoked. The objective is to maximize the contribution margin gained from accepted orders minus holding and backlog costs, which depend on the actual finish dates of the orders.

The problem is formulated as a stochastic dynamic program to show differences to other problems considered already in literature. In particular, this article is the first to consider online order acceptance and order release decisions in a multi-stage production system. For instances of realistic size, the stochastic dynamic program cannot be solved due to the high dimensionality of the resulting state space. Hence, a bid price-based revenue management approach is developed. Here, the multi-stage production system is modeled as a linear program. Bid prices are derived from this linear program and aggregated to opportunity costs for starting an order in a given planning period. An order is accepted if there exists a feasible release date for the order and the expected profit from this order exceeds the opportunity costs of releasing the order in this period. The order release decisions are made via a linear program that ensures feasibility of the schedule and uses the opportunity costs in the objective function.

A numerical study shows the good performance of the developed approach compared to an ex-post optimal solution, which has knowledge about all incoming demand upfront. It is shown that aggregating the production system to a single-stage system leads to unsatisfactory results in complex production systems. Making online order release decisions significantly increases the profit gained compared to fixing the release date immediately upon arrival of an order.

[2]Guhlich, H., Fleischmann, M., Mönch, L., and Stolletz, R. (2015a). Revenue management in a multi-stage make-to-order production system. Working paper, University of Mannheim.

The third article (joined work with Moritz Fleischmann, Lars Mönch, and Raik Stolletz)[3] considers a multi-stage production system, which is subject to stochastic influences. That is, both the arrival process of the demand and the production process are stochastic. Order acceptance decisions are made immediately upon arrival of each order, while order release decisions are made only at the beginning of each planning period. Partial order acceptance and partial order releases are possible. The objective is to maximize the contribution margin gained from accepted orders minus finished good inventory costs, work-in-process holding costs, and backlog costs.

This article is the first to consider order acceptance and order release decisions for stochastic demand in a production system with stochastic influences. It combines ideas of bid price-based revenue management with production planning approaches based on clearing functions. More precisely, clearing functions are used to model the non-linear dependency of workload and lead times in stochastic production systems. The clearing functions are approximated by piece-wise linear functions to describe the production process with a linear program. This linear program is used to compute bid prices, which are aggregated to opportunity costs for using a machine group in a given planning period. The order acceptance and order release decisions are made via the same linear program using the opportunity costs in the objective function.

A numerical study shows the benefit of this novel approach compared to approaches that cope with variability in the production system by modeling fixed, workload-independent waiting times, which is common practice in order release planning. The results are derived in basic production systems and confirmed by the study of a model of a real-world semiconductor back-end facility.

[3]Guhlich, H., Fleischmann, M., Mönch, L., and Stolletz, R. (2016). Revenue management in stochastic production systems: A clearing function-based bid price approach. Working paper, University of Mannheim.

2 Revenue management approach to due date quoting and scheduling in an assemble-to-order production system

Co-authors:

Moritz Fleischmann
Chair of Logistics and Supply Chain Management, Business School, University of Mannheim, Germany

Raik Stolletz
Chair of Production Management, Business School, University of Mannheim, Germany

Published in:

OR Spectrum, Volume 37(4) (2015), pages 951-982, DOI: 10.1007/s00291-015-0401-3, reprinted with permission of Springer

Abstract:

In this paper, we consider demand management decisions for an assemble-to-order production system in which both the availability of intermediate material and assembly capacity are limited. For each incoming order, the manufacturer must decide whether to accept it and what due date to quote for an accepted order. The actual assembly dates are still subject to change after these decisions, and a production schedule must be maintained to guarantee that the quoted due dates are met. Therefore, the decisions on accepting orders and quoting due dates must be made with incomplete knowledge of the actual resources used to fulfill the orders. To address these factors, we model this situation and develop a novel revenue management approach using bid prices. An extensive numerical study demonstrates the good performance of the proposed approach in comparison with benchmark algorithms and an ex-post optimal solution applied over a wide range of different supply and demand

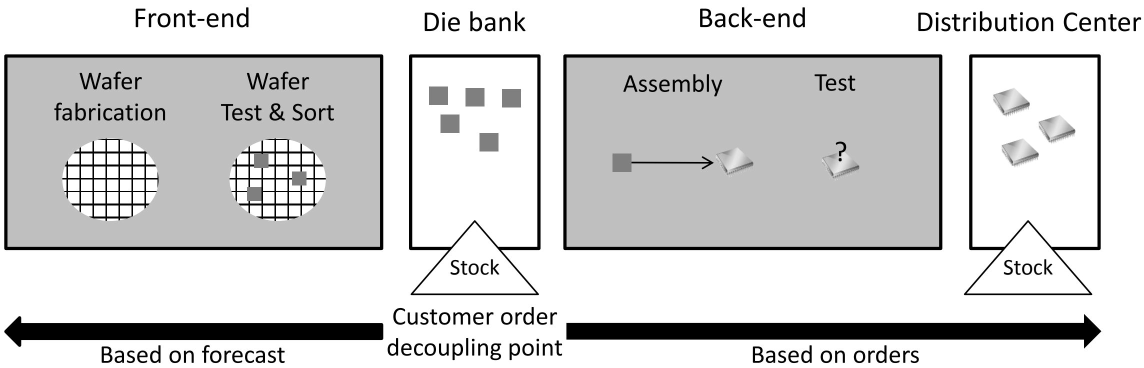


Figure 2.1: Supply chain structure of a semiconductor manufacturer

scenarios. Our results suggest that the consideration of assembly capacity constraints is more vital than the consideration of intermediate material constraints in our test cases.

2.1 Introduction

In an assemble-to-order (ATO) production system, intermediate materials are assembled into end products. Intermediate materials are held in stock because they often have long replenishment lead times. End products are assembled only for customer orders on hand.

For example, ATO production systems are found in high-tech industries, where capital equipment costs are high and lead times for intermediate materials are long. Therefore, in the short-term, it is impossible to adjust capacity levels. The demand is typically fluctuating and uncertain. The production is order-driven due to a high risk of obsolescence or customer-specific configurations of the end products.

In companies with high capital equipment costs, the capacities are tightly planned because unused capacity is highly expensive. Therefore, if these companies face uncertain and fluctuating demand, resources may be insufficient to fulfill all incoming demand during peak phases. As customers are heterogeneous (i.e., they differ in their strategic importance, time-sensitivity, willingness to pay, etc.), the question arises as to which customers to serve using the scarce resources.

As an example, Fig. 2.1 depicts the supply chain of a semiconductor manufacturer (cf. Brown et al. (2000)). In the front-end production, wafers are fabricated, tested, and sorted. The finished wafers are stored in a *die bank*, which also marks the customer order decoupling point. In the back-end process, at an assembly-and-test facility, the wafers are taken from the die bank, processed, and assembled to finished products, according to customer specific requirements. The paper at hand considers the back-end manufacturing because only here demand fulfillment decisions are made. Assembly capacity and wafer inventory are limiting resources. Demand in this environment is often uncertain and fluctuating, with peak phases in which demand exceeds supply. The lead times of wafers are long, compared to the assembly time for finished products. In peak phases, available resources must be allocated by quoting different lead times. If due dates are set for customers, there is still freedom in how to schedule order assembly. This must be done in an online fashion because full information on future orders is not available. For some of the customers it is very important to get their chips in time (within the peak phase) because their own production requires them. Other customers are less time-critical. Therefore, some customers are willing to pay a premium for short lead times. Also, quoting long lead times bears the risk of losing the order to competitors. This leads to different profit margins from different customers, generated by the same products. This scenario represents a typical revenue management problem, which we analyze in this paper.

Traditional revenue management approaches focus on the decision of which orders to accept (see, e.g., Talluri and van Ryzin (1999), Spengler et al. (2007)). However, in practice, the company typically must decide on the delivery time for each customer as well. Apart from giving discounts, failure to meet the customer's preferred due date can also lead to a loss of goodwill of the affected customer. Ultimately, the company might even lose this customer to competitors, which is why it is important to meet the due date preferences of customers with high value to the company.

The actual assembly dates of accepted orders are still flexible up to a certain point. If preemption is not allowed, it might be beneficial to delay assembly of an order, even though assembly capacity is available, in anticipation of a rush order with higher value.

In this paper, we consider demand management decisions for a manufacturer using an ATO production system and facing stochastic demand. To maximize profit, the manufacturer decides which orders to accept and what due dates to quote for the accepted orders. A production schedule is maintained to guarantee that the quoted due dates are met. The availabilities of both assembly capacity and intermediate material are limited.

The main contributions of this paper are as follows:

- We introduce a novel decision problem and model it as a stochastic dynamic program (SDP).

- We develop a heuristic revenue management approach (ATO-RM) using bid prices to address the described problem. In particular, we introduce a scheduling approach that captures possible future orders via bid prices. Using an extensive numerical study, we show that the approach works well by comparing it to benchmark algorithms and an ex-post optimal solution.
- A sensitivity analysis also shows that the consideration of constraints relative to assembly capacity appears more important to obtaining good results than the consideration of constraints on intermediate material in our test cases.

This paper is organized as follows. After reviewing the literature in Section 2.2, we describe the assumptions of our model in Section 2.3. Next, we present an SDP modeling the problem exactly in Section 2.4. Next, we present our approach for quoting due dates and performing the scheduling in Section 2.5. The corresponding calculations require bid prices, which we derive in Section 2.6. A numerical study demonstrates the performance of our approach in Section 2.7. In Section 2.8, we conclude and discuss extensions to cope with modified assumptions.

2.2 Literature review

In this section, we review the literature closely related to our research. First, we survey the stream of literature that considers decisions of production planning in ATO manufacturing in Section 2.2.1. A growing body of literature exists in the area of revenue management in manufacturing, which is presented in Section 2.2.2. Finally, we examine the emerging literature on revenue management of flexible products in Section 2.2.3. For each of the presented literature streams, we first introduce the relevant papers and then explain the differences to our work.

2.2.1 Production planning in ATO manufacturing

Song and Zipkin (2003) give an overview of supply chain operations in ATO manufacturing. Kolisch (2001) describes a hierarchical planning approach used to model the decisions of an ATO manufacturer. An order selection process is placed on the first hierarchy level to decide on accepting orders and quoting due dates. The next hierarchy levels are manufacturing planning and operations scheduling. The actual production planning steps are modeled in much greater detail than in our research, but the previously described models are deterministic, and thus, order acceptance and due date quoting decisions are not made online as is the case in the present paper.

Benjaafar and ElHafsi (2006), ElHafsi (2009), and Cheng et al. (2011) contribute to another stream of literature that addresses production and inventory control in ATO manufacturing. Their models include various customer classes with different costs for lost sales, leading to an order selection problem. However, in contrast to the model presented in this work, due date quoting decisions are not taken into account. Only one end product that requires several intermediate materials for production is considered. The time required for assembly is assumed to be negligible, and thus these authors do not model assembly capacity; however, they decide when to produce the intermediate products. The production time for intermediate materials is stochastic, and orders cannot be backlogged in their setting.

Standard available-to-promise (ATP) approaches allocate available resources to existing customer orders. Their main purpose is to determine whether there are enough resources available to accept an order. Advanced available-to-promise approaches (cf. Chen et al. (2001), Chen et al. (2002), Pibernik (2005)) decide about a due date and the quantity of products promised. They maximize total profit, typically including revenues for accepted orders, holding/backlog costs, and lost-sales costs. Several authors adapt existing MIP approaches to specific production environments (e.g. Zhao et al. (2005), Tsai and Wang (2009), and Lin et al. (2010)).

In this literature stream, ATO production with material and capacity constraints are considered as in this paper. They also decide about acceptance of orders and due date quoting as well as on a production schedule. However, the difference is that in their setting, orders are deterministic, while we consider a stochastic online problem. Therefore, the proposed approaches are not applicable in our setting. Furthermore, in ATP, orders are rejected only if not enough resources are available, while in revenue management, due to the heterogeneity of customers, orders are also rejected in anticipation of more valuable orders.

2.2.2 Revenue management in manufacturing

Revenue management approaches originally focused on applications in service industries such as the airline or hotel business, see Talluri and Van Ryzin (2004). The literature concerned with revenue management in manufacturing can be categorized by the underlying production principle.

Revenue management in make-to-stock production systems

In the context of make-to-stock (MTS) revenue management, only storable resources are considered, whereas one of the main characteristics of the model presented in the present paper is the combination of storable and non-storable resources, see e.g. Meyr (2009) and Quante et al. (2009). Liu (2005) discusses two different MTS revenue management models, and one model where one order stream is MTS and one order stream is make-to-order (MTO).

Revenue management in make-to-order production systems

In MTO revenue management, one important literature stream applies revenue management ideas in the iron and steel industry in which unique orders arrive over time. Spengler et al. (2007) model the decision of which orders to accept in a setting with multiple resources. However, the production time does not exceed one planning period as opposed to the setting in the present paper in which resources are required over multiple planning periods. Their decisions are based on bid prices, which are computed via a multi-dimensional knapsack problem formulation. Spengler et al. (2008) extend a one-period bid price approach to multiple periods and allow for shifting demand between periods. An order is accepted if its revenue exceeds the sum of the bid price for production capacity in this period and the holding costs. Hintsches et al. (2010) base their work on the general framework of Spengler et al. (2007) for order acceptance. They use a bid price approach based on a linear program that models the scheduling decisions. Volling et al. (2012) decide about the use of one bottleneck resource. They update bid prices using neural networks.

Barut and Sridharan (2005) develop a heuristic approach to compute the capacity contingents for several customer classes to decide which orders to accept in an order-driven production system. Gallien et al. (2004) describe a model for admission control in a single server queue with preemption allowed at no cost. This group makes their decisions by computing a three-dimensional acceptance region consisting of price, quantity, and lead time. Kuhn and Defregger (2005) present a revenue management approach for a MTO manufacturer with limited end-product inventory. In each period, the company decides whether the incoming order is accepted. In this case, it must be determined to what extent the order is satisfied from stock. Additionally, in each period, the company must decide if new end products should be produced and put into stock. Kniker and Burman (2001) use a

Markov chain approach for a setting, where orders arrive with a fixed due date and orders are scheduled using the FCFS principle.

The papers in this literature stream consider only production capacity, the availability of raw material is disregarded. This is a main difference to the setting in the paper at hand, where both resources are scarce. Storable resources must be treated differently because of their special structure. In contrast to capacity, they are still available if not used in the current period. Additionally, they are replenished in pre-specified periods, while capacity is available only at a given point in time. Also, these papers assume that production of an order is done within one period, while the lead times considered in this paper span multiple planning periods. In addition, the combination of due date quoting and the actual scheduling of orders is a contribution of the present paper.

Revenue management in assemble-to-order production systems

Harris and Pinder (1995) are the first to mention revenue management in an ATO context. They validate that the requirements for successfully applied revenue management are fulfilled in ATO manufacturing. This paper also describes a pricing and capacity allocation approach for a simplified model. Gao et al. (2012) model an available-to-promise assembly system. They use pseudo-orders to model uncertainty in the orders and develop a Markov chain model to obtain insights into optimal order acceptance decisions. This model also includes inventory and capacity constraints, but the raw material is not required to be available at the time of the production start. Accepted orders are processed in a first-come-first-served manner. The existing approaches in ATO revenue management rely on dynamic programming. Order processing, including due date quoting and scheduling, is not modeled because it would render these approaches intractable due to the resulting large state space.

2.2.3 Revenue management of flexible products

Gallego and Phillips (2004) introduce the concept of revenue management of flexible products (FP-RM). Their model includes both specific and flexible products. Although the capacity requirements are fixed for specific products, every flexible product has a set of possible execution modes. Each execution mode can lead to a different level of resource consumption and profit margin. To save capacity for the higher-margin specific products that are yet to come, orders can be rejected and the right execution modes for the flexible products must be chosen. At a fixed point in time, the so-called notification date, the execution modes of all flexible products must be fixed. After that, only demand for specific products occurs. Gallego and Phillips (2004) look at a two-period/two-product case, and due to the small size of the problem, they are able to compute optimal booking limits using dynamic programming.

Gallego et al. (2004) consider two different settings, namely independent, exogenously generated demand, and demand driven by a customer choice model. In our setting, independent, exogenous demand is assumed. For this setting, they extend the approach of Gallego and Phillips (2004) to a network setting with an arbitrary number of periods. They show that their proposed deterministic linear program (DLP) approach is an asymptotically optimal approximation of the stochastic optimization problem. In the network setting, the size of the dynamic program vastly increases such that they introduce a bid price approach to solve the problem heuristically. For the customer choice model, they also find an asymptotically optimal DLP approach.

Petrick et al. (2010) use a similar bid price approach as is presented in Gallego et al. (2004). They derive several dynamic control mechanisms that differ in their degree of flexibility.

Petrick et al. (2012) further extend the FP-RM concept by introducing a model formulation that allows for reallocation of the execution mode for flexible products and a more general choice of the notification date. They formulate an SDP and motivate their bid price approach by showing that it is consistent with an optimal solution of the primal linear program.

Gönsch et al. (2014) define the value of flexibility as the value for the manufacturer of being able to choose the optimal execution mode at the end of the booking horizon instead of allocating it immediately at arrival. Existing DLP models do not consider that flexible products can be re-arranged later and thus underestimate the value of flexibility. The authors propose an approach to incorporate the right value of flexibility by adapting revenues dynamically over time.

FP-RM has some analogies compared to a manufacturing setting. The different assembly-times can be interpreted as possible execution modes. Accordingly, different resource consumption and profit margins can be modeled. However, the time structure of our model is different, which makes the described approaches not applicable in the setting of the present paper. In the papers in this stream, service is provided after all flexible orders have arrived, but these events overlap in our model. Hence, we must make scheduling decisions, which cannot be revoked later, i.e., they must be made with less information available compared to the FP-RM setting.

Additionally, the execution modes are time-dependent in our setting. Unused assembly capacity is lost. Therefore, some execution modes are no longer available as time elapses. This must be considered to make appropriate decisions. The introduction of intermediate materials means that some execution modes only become available over time. This cannot be modeled using existing FP-RM approaches because here the service provision is after the decision period, and the final decision of the allocation to the execution modes is done at the end of the planning horizon. Furthermore, the available execution modes are limited by due date quoting decisions that we make.

2.3 Problem formulation

In our problem setting, orders arrive over time according to a known probability distribution. The task is to decide on the acceptance and quoted due dates for each incoming order. A feasible schedule must be maintained to guarantee that the quoted due dates are met. Assembly capacity and intermediate materials are the limiting resources.

In this section, we describe the relevant decisions and basic assumptions of our model, the assembly process, the properties of the orders, and the order processing. Table 2.1 summarizes the notation.

The considered planning horizon is discretized into T *planning periods* ($t = 1, \ldots, T$). We further divide each planning period t in S *micro-periods* ($s = 1, \ldots, S$) such that in each of these micro-periods (t, s) at most one order arrives. Incoming orders are denoted by $d = 1, \ldots, D$.

Assumption 2.3.1 (Decisions) *The company makes the following decisions:*

1. Due date quoting: *For each incoming order, decide the quoted due date δ_d (or reject it).*
2. Scheduling: *For each accepted order $d \in Y$, decide what fraction $x_{dt} \geq 0$ of the order d starts assembly in the current planning period t.*

Y denotes the set of orders that are accepted but for which assembly has not yet started for the whole order at the beginning of the current planning period t. That is, each member of Y is a tuple $(d, \delta_d^{fix}, x_{dt}^{cum})$ that describes the order d, the quoted due date δ_d^{fix}, and $x_{dt}^{cum} := \sum_{t'=1}^{t-1} x_{dt'}$, the cumulative fraction of the order for which assembly has already started up to t.

Indices:	
$t = 1, \ldots, T$	Planning periods of the planning horizon
$s = 1, \ldots, S$	Micro-period s in a planning period
$d = 1, \ldots, D$	Orders
$m = 1, \ldots, M$	Intermediate material types
Parameters:	
$contr_d$	Contribution margin of d
$pref_d$	Preferred due date for d
use_{dm}	Amount of intermediate material m required for d
arr_d	Arrival planning period of d
$lead_d$	Required number of consecutive planning periods of assembly capacity for d
c_d^H / c_d^B	Holding/Backlog cost rate (per planning period) for d
hc_{dt}	Holding costs for d if assembly starts in t
bc_{dt}	Backlog costs for d if t is quoted due date
δ_d^{fix}	Quoted due date for d if d is accepted order
a_t	Available assembly capacity in t
$r_t = (r_{1t}, \ldots, r_{Mt})$	Replenishment of intermediate materials in t
$cap(x_t)$	Assembly capacity used for assembly x_t
$mat(x_t)$	Intermediate materials used for assembly x_t
p_{dts}/p_{0ts}	Probability that order d / no order arrives in (t, s)
State variables:	
a'	Vector of remaining assembly capacity
$ATP_{mt} \geq 0$	Available-to-promise quantity of m at the end of t
Y	Set of accepted orders for which assembly is not yet started for the whole order
x_{dt}^{cum}	Cumulative fraction of d for which assembly is started up to t
Decision variables:	
$x_t = (x_{1t}, \ldots, x_{Dt})$	Vector indicating for what fraction of each order the assembly starts in t
δ_d	Quoted due date for d

Table 2.1: Notation

Due date quoting must be made instantly after order arrival, i.e. in each micro-period. Scheduling decisions are made at the beginning of every planning period.

These decisions are based on knowledge of the available resources, the previously accepted orders, their assembly status, and the probability distribution for future incoming orders.

Assumption 2.3.2 (Assembly process) *Assembly is modeled as a deterministic one-stage process. Setup-times are not considered. The constrained resources are assembly capacity* a_t *and the availability of intermediate materials* $m = 1, \ldots, M$ *in each planning period* t*. Each order* d *requires one unit of assembly capacity in* $lead_d$ *consecutive planning periods. Additionally, at the start of the assembly,* use_{dm} *units of intermediate material* m *are required.*

Each order consists of multiple identical products so that it is possible to start assembly for fractions of an order (i.e., only for a fraction of the required products) in different planning periods and to deliver them as soon as their assembly is finished. If the assembly process for (a fraction of) an order has been started, it has to be processed for $lead_d$ *planning periods without interruption due to technical reasons. In those periods, only the respective fraction of one capacity unit is required. That is, splitting up an order is allowed but preemption is not.*

Note that the lead time is independent of the lot size because the processing time of each single product is assumed to be constant. That is, assembling additional products requires more capacity but does not take more time (if sufficient capacity is available). Actual detailed shop floor scheduling is not the goal of the proposed model. Therefore, we model a simplified assembly process. In revenue management approaches, this is a common assumption (cf. e.g. Hintsches et al. (2010), Gao et al. (2012)). The main decisions are whether to accept an incoming order and the choice of a quoted due date. For these decisions, the remaining assembly capacity must be estimated.

Assumption 2.3.3 (Resources) *A short-term planning horizon is considered. Within this time horizon, the company cannot extend the maximum capacity levels. A replenishment amount* r_{mt} *is exogenously given for each intermediate material* m *in each planning period* t*.*

The used capacities resulting from the scheduling decision $x_t = (x_{1t}, \ldots, x_{Dt})$ *are defined as follows. The vector of assembly capacity consumption of orders starting in* *planning period t according to x_t is $cap(x_t)$. In particular, the i–th element of $cap(x_t)$ describes the capacity consumption in planning period $t+i$ for $i = 0, \ldots, T-t$ and is defined as $(\sum_{d \text{ with } lead_d \geq i} x_{dt})$. The intermediate materials required for the assembly starts are described by $mat(x_t)$, which is a vector consisting of elements $\sum_d use_{dm} \cdot x_{dt}$ for $m = 1, \ldots, M$.*

Let a' denote the vector of remaining assembly capacities in the planning periods to come. Capacities in the past can be disregarded. Let ATP denote the available amount of each intermediate material at the end of the current planning period.

This assumption reflects the fact that capacity adjustments are mid- to long-term decisions and intermediate materials often have long replenishment lead times, whereas our model is used for short-term decisions. It follows that a_t and r_{mt} are not decisions in the model presented in this work.

Assumption 2.3.4 (Orders) *In each micro-period, orders arrive according to a known probability distribution. We denote by p_{dts} the probability that order d arrives in micro-period (t, s). p_{0ts} is the probability that no order arrives in this micro-period. Each order d has the following characteristics:*

- *Arrival planning period (arr_d)*
- *Preferred due date ($pref_d$)*
- *Amount of required intermediate material (use_{dm}) for $m = 1, \ldots, M$*
- *Required number of consecutive planning periods of one unit of the assembly capacity ($lead_d$)*
- *Contribution margin ($contr_d$) consisting of revenue less variable production costs*
- *Holding cost rate (c_d^H)*
- *Backlog cost rate (c_d^B)*

Future demand is stochastic but forecasts are available and are represented by probability distributions for the incoming orders. The contribution margin does not include backlog and holding costs.

Assumption 2.3.5 (Time structure within a planning period) *At the beginning of a planning period, the assembly processes of the end products are finished. Therefore, products*

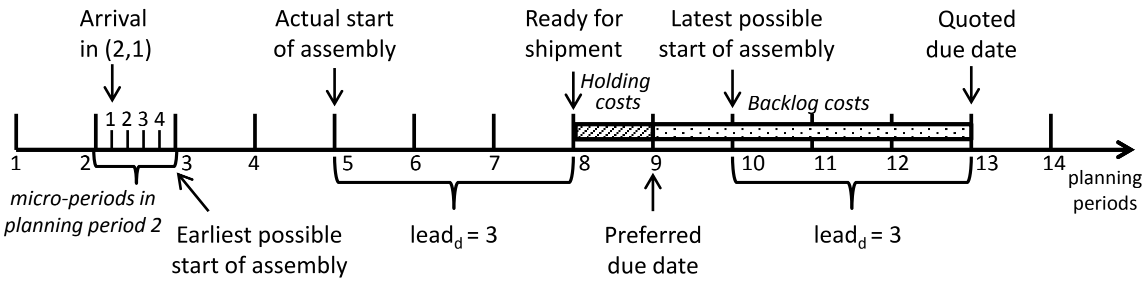


Figure 2.2: Example of order processing

can be delivered in the same planning period in which they were finished. Intermediate material replenishment occurs before the assembly process such that the replenished material can be used within the same planning period. Orders can start assembly at the earliest in the next planning period after their arrival.

The next assumption describes the order processing steps in our model. For better understanding, Fig. 2.2 depicts an example of the processing of an order. In the following, the values for the example are given in parentheses.

Assumption 2.3.6 (Order processing) *The customer requests delivery on a preferred due date $pref_d$ $(= 9)$. Upon order arrival (in micro-period (2,1)), the company immediately either quotes a due date δ_d $(= 13)$ to the customer, which must be met, or rejects the order. The decision about the actual delivery date is not made instantly. The company can begin the assembly in any planning period after the arrival such that the quoted due date is met (in planning periods 3 to 10). The quoted due date is always accepted by the customer. If the quoted due date is beyond the preferred due date, backlog costs occur (for 4 planning periods). Finishing assembly before the preferred due date (in planning period 8) results in holding costs up to the preferred due date (for 1 planning period). Note that under these assumptions, both holding and backlog costs can occur for a given order if the quoted due date is after the preferred due date and the order is ready before the preferred due date.*

Orders can only be fully accepted. However, recall that it is possible to fulfill the order via partial deliveries over different planning periods (cf. Assumption 2.3.2).

Due dates must be quoted online because customers request an immediate answer. Once the company has committed to a due date, this due date must be met. If t is the *quoted due date* for order d, the resulting backlog costs are

$$bc_{dt} := (t - pref_d)^+ \cdot c_d^B, \tag{2.1}$$

where c_d^B is the backlog cost rate. [1]

If the assembly process is finished before the preferred due date, holding costs are incurred because early delivery is not desired by the customer, and customers are willing to get their products on their preferred due date independent of the quoted due date. If t is the actual *start of assembly* for order d, the resulting holding costs are

$$hc_{dt} := (pref_d - (t + lead_d))^+ \cdot c_d^H, \tag{2.2}$$

where c_d^H is the holding cost rate. It follows that it is never beneficial to quote a due date that is earlier than the preferred due date.

Assumption 2.3.7 (Pricing) *Prices are fixed in the short-term.*

We assume that prices are given exogenously. Pricing decisions in a similar setting as presented in the current paper are considered in Charnsirisakskul et al. (2006).

2.4 Stochastic dynamic program

In this section, we model the above described problem as a stochastic dynamic program (SDP). We define the state space, decisions, and transitions, and present the Bellman equation for the considered problem. The feasibility of a state is checked in a linear program, which we also describe in this section. Additional notation is summarized in Table 2.2.

State space

The state of the system is described by $\xi = (a', ATP, Y)$, the available resources and the orders that are already accepted but which are not yet completely assembled.

We define $\mathcal{Z}$ to be the set of feasible states. A state is feasible if each accepted order can be produced within its quoted due date. This can be checked using a linear program, which is defined at the end of this section.

[1] In the present paper, the notation x^+ is used as short for $\max\{0, x\}$.

Sets:	
$X_{t\xi}$	Set of feasible assembly start decisions in planning period t in state ξ
$\mathcal{S}_{d\xi}$	Set of feasible due dates for an order d in state ξ
$\mathcal{Z}$	Set of all feasible states
Profit function:	
$V_{ts}(\xi)$	Maximum expected marginal profit-to-go in state ξ in planning period t and micro-period (t, s), respectively
State variable:	
$\xi = (a', ATP, Y)$	State of the system given by remaining available assembly capacity a', ATP quantity ATP, and the set of fractions of orders to be scheduled Y

Table 2.2: Additional notation for SDP

Decisions and transitions

Due date quoting decisions must be made immediately after order arrival and scheduling decisions are made only at the beginning of a planning period. Therefore, different decisions and transitions must be defined for both micro and planning periods.

In each *micro*-period, we must decide about acceptance of the arriving order and quote a due date for accepted orders (cf. Assumption 2.3.1). If the order is rejected (or no order arrives), only s is increased by 1. Otherwise, the accepted order d with its quoted due date δ_d is added to the set of accepted orders, denoted by $Y \cup (d, \delta_d, 0)$. Note that at this point no resources are used. The optimal due date must be chosen from the set describing the due dates with which the resulting state remains feasible:

$$\mathcal{S}_{d\xi} := \{\delta_d \in \{pref_d, \dots, T\} \mid (a', ATP, Y \cup (d, \delta_d, 0)) \in \mathcal{Z}\}. \tag{2.3}$$

At the beginning of each *planning* period, we must decide what fraction of each accepted order to start assembling. This corresponds with the *scheduling* decision from Assumption 2.3.1. Based on this decision, the resource availability for the upcoming planning periods needs to be reduced.

Assembly can only be started for orders that are accepted but for which assembly has not yet completely started. New assembly starts must not exceed the order volume, and capacities must suffice for the planned assembly. Therefore, the set of possible assembly start decisions in planning period t is described by

$$X_{t\xi} := \left\{ x_t \in [0,1]^D \;\middle|\; \begin{array}{l} x_{dt} = 0 \text{ for } d \notin Y, \\ x_{dt} + x_{dt}^{cum} \leq 1 \text{ for } d \in Y, \\ (a' - cap(x_t), ATP + r_t - mat(x_t), Y - x_t) \in \mathcal{Z} \end{array} \right\} \tag{2.4}$$

Here, the state transition $Y - x_t$ describes the update of the set of fractions of orders that are still to be scheduled. In particular, the cumulative assembly starts $x_t \in X_{\xi_t}$ are adjusted by $x_{dt+1}^{cum} := x_{dt}^{cum} + x_{dt}$ and each order d with $x_{dt+1}^{cum} = 1$ is removed from the set Y. The remaining available capacity in planning period $t-1$ can be dropped from the state space.

Profits

Let $V_{ts}(\xi)$ be the maximum expected marginal profit-to-go of state ξ at the beginning of planning period t if $s = 0$ and in micro-period (t,s) for $s = 1, \ldots, S$, respectively. Additionally, we define $V_{t,S+1}(\xi) := V_{t+1,0}(\xi)$. Then, V_{ts} can be computed recursively via the Bellman equation for $\xi \in \mathcal{Z}$ as follows:

$$\begin{aligned} &V_{ts}(\xi) = V_{ts}(a', ATP, Y) = \\ &\begin{cases} \sum_d p_{dts} \cdot \max\Big\{ \max\limits_{\delta_d \in \mathcal{S}_{d\xi}} \big\{ V_{t,s+1}(a', ATP, Y \cup (d, \delta_d, 0)) + contr_d - bc_{d\delta_d} \big\}, \\ V_{t,s+1}(\xi) \Big\} + p_{0ts} \cdot V_{t,s+1}(\xi) & \text{if } 1 \leq s \leq S \\ \max\limits_{x_t \in X_{t\xi}} \Big\{ V_{t,s+1}(a' - cap_t(x_t), ATP + r_t - mat(x_t), Y - x_t) \\ -\sum_d x_{dt} \cdot hc_{dt}) \Big\} & \text{if } s = 0 \end{cases} \end{aligned} \tag{2.5}$$

with boundary conditions

$$V_{T,S+1}(\xi) = 0 \quad \text{if } \xi \in \mathcal{Z}. \tag{2.6}$$

It is insightful to compare SDP (2.5) to the SDP for FP-RM presented in Petrick et al. (2012). Their model formulation additionally considers specific orders, and the scheduling decision is performed only once, at the notification date.

In (2.5), we make irrevocable scheduling decisions in every planning period. Therefore, the cumulative amount of assembled products for an order (x_{dt}^{cum}) must be included in the state space. Additionally, we make due date quoting decisions that limit the available assembly start periods for an order. Thus, the quoted due date (δ_d^{fix}) is part of the state space for each accepted order d. In our model, resources are modeled differently due to the underlying time structure. Assembly capacity of past planning periods is not available anymore, and new intermediate material replenishments become available in each planning period ($ATP_t + r_t$). In FP-RM terms that means that the availability of execution modes is time-dependent and can be influenced by due date quoting. In Petrick et al. (2012), execution modes are always available up to the notification date.

Feasibility check

To check *feasibility*, that is if a state is in $\mathcal{Z}$, we test whether a feasible schedule exists by solving a linear program (LP). We use this LP throughout the paper as feasibility must also be guaranteed in the due date quoting and scheduling decisions in Section 2.5. Additionally, this LP is the basis of the bid price computation presented in Section 2.6.

To test if order d arriving in micro-period (t^{fix}, s) can be feasibly accepted with the quoted due date δ_d^{fix}, we set $Y := Y \cup (d, \delta_d^{fix}, 0)$ and test the feasibility of the state (a', ATP, Y) at the beginning of the next planning period ($t^{fix} + 1$). We model the decisions by x_{dt}, which indicates for what fraction of d the assembly begins in planning period t. Variables x_{dt} are defined only for orders $d \in Y$ and in planning periods t with $max(t^{fix} + 1, arr_d + 1) \leq t \leq T - lead_d$ to prevent that assembly for an order begins before its arrival or that decisions are taken for previous planning periods. Additionally, this ensures that assembly is finished at the end of the planning horizon. State (a', ATP, Y) is feasible at the beginning of planning period $t^{fix} + 1$ if the following inequalities are satisfied: [2]

[2]We follow the convention that undefined variables are equal to 0.

$$\sum_{d \in Y} \sum_{i=0}^{lead_d - 1} x_{d(t-i)} \leq a'_t \quad \text{for } t = t^{fix} + 1, \ldots, T \tag{2.7}$$

$$ATP_{mt} - ATP_{m(t-1)} + \sum_{d \in Y} use_{dm} \cdot x_{dt} = r_{mt} \quad \begin{array}{c}\text{for } t = t^{fix} + 1, \ldots, T; \\ m = 1, \ldots, M\end{array} \tag{2.8}$$

$$\sum_{t=arr_d+1}^{\delta_d^{fix} - lead_d} x_{dt} + x_{dt^{fix}}^{cum} = 1 \quad \text{for } d \in Y \tag{2.9}$$

$$x_{dt} \geq 0 \quad \text{for } d \in Y \; t = t^{fix} + 1, \ldots, T \tag{2.10}$$

The remaining available assembly capacity cannot be exceeded in any planning period. Therefore, the capacity consumption in planning period t takes into account all orders d, which started (partial) assembly in the $lead_d$ planning periods before (cf. (2.7)). Note that orders that started assembly before the current planning period are not defined in the model and the assembly capacity used by them is already considered in the remaining available assembly capacity a'. Balance equation (2.8) models the inventory of intermediate materials and must apply for each planning period and each intermediate material type. In this work, ATP_{mt} is the available-to-promise quantity of type m at the end of planning period t. By constraint (2.9), each accepted order must be scheduled such that the quoted due date is met. Fractions of the orders that started assembly before the current planning period are cumulated in $x_{dt^{fix}}^{cum}$. Note that due to this equation, the entire order must be processed. Additionally, non-negativity constraints (2.10) are required.

We note that the rather general formulation of the LP model in Petrick et al. (2012) can also be adapted to check feasibility in our setting. This is because in the LP formulation (2.7)-(2.10) no future demand is considered and therefore, the different time structure in comparison to Petrick et al. (2012) is no longer relevant. However, to adapt their model, a very large number of execution modes is required. Specifically, for each combination of possible assembly start and used intermediate material replenishment, an additional execution mode must be created. Hence, model formulation (2.7)–(2.10) is more efficient in the considered setting.

```
Compute bid prices;
for t = 1 to T do
    Schedule accepted orders (using approximated opportunity costs);
    Update bid prices;
    foreach order d arriving in t do
        Quote due date for d or reject d (using approximated opportunity costs);
    end
end
```

Algorithm 1: General approach

2.5 Revenue management approach

Due to the curse of dimensionality, it is computationally intractable to solve the described SDP (2.5) exactly even for rather small problem instances (cf. e.g. Petrick et al. (2012)). Therefore, we need a heuristic approach to make the decisions stated in Assumption 2.3.1. We describe this approach in this section. We base the decisions on opportunity cost estimates for the used resources. As is common in the literature, bid prices are used as an estimate for the real opportunity costs; see, e.g., Simpson (1989), Williamson (1992) or Talluri and Van Ryzin (2004). We use the estimated opportunity costs in this section but postpone their computation until Section 2.6. We start by describing how to decide on due date quoting. Next, we explain how to schedule the accepted orders. The general approach is summarized in Algorithm 1.

2.5.1 Due date quoting

For each incoming order, we search for a *feasible* due date that maximizes the *profitability* of the order. The decision on accepting orders is integrated into the due date quoting decision. Profitable orders are accepted with the respective due date, and non-profitable orders are rejected. In the following, the profitability metric of an order is defined. The *feasibility* of a schedule is checked using LP (2.7)–(2.10).

We measure the *profitability* of an order d with quoted due date δ_d by

$$(contr_d - bc_{d\delta_d}) - opp^{due}_{d,\delta_d}, \tag{2.11}$$

which is the difference between the immediate marginal profit earned and opp^{due}_{d,δ_d}, the opportunity costs for accepting d with due date δ_d. Note that holding costs, which depend on the actual assembly date, are included in opp^{due}_{d,δ_d} as will be explained in Section 2.6.2.

2.5.2 Scheduling

In this section, we explain how to schedule the accepted orders. At the beginning of each planning period, the orders that begin their assembly in this period must be determined. Therefore, we set up a linear program that includes all accepted orders. As the schedule must be feasible, inequalities (2.7)–(2.10) are included. Next, we need to define a suitable objective function. The decisions that this linear program suggests for the current planning period are taken, and decisions for later periods are discarded.

As the orders are already accepted with given due dates, cost wise, scheduling can only influence the resulting holding costs. However, if the schedule is optimized only with respect to the holding costs, orders tend to begin their assembly late, which leads to unused capacities in the upcoming planning periods.

A better measure for the quality of a solution can be found by also taking into account the incoming future demand with the use of opportunity costs, which are used to minimize the usage of the most profitable resources. Let opp^{res}_{dt} denote the opportunity costs for order d with given actual start of assembly in planning period t. Next, we define the objective function as follows:

$$\sum_{d=1}^{D} \sum_{t=arr_d+1}^{\delta^{fix}_d - lead_d} (opp^{res}_{dt} + hc_{dt}) \cdot x_{dt}. \tag{2.12}$$

Minimizing (2.12) with respect to constraints (2.7)–(2.10) gives the preliminary schedule. The decisions made in the current planning period t are fixed from this point on and the remaining capacities must be adapted: $a' := a' - cap_t(x_t)$.

2.6 Opportunity costs

In this section, we explain how to derive the opportunity cost estimates used in the previous section. First, we show how to compute bid prices for assembly capacity (bp^{cap}_t) and intermediate material (bp^{mat}_{mt}). Then, we approximate opportunity costs of accepting an order with given start of assembly (opp^{res}_{dt}) and with given due date $(opp^{due}_{d\delta_d})$, respectively. Finally, we update bid prices if new information is available.

2.6.1 Bid price computation

In this section, we describe how to compute bid prices for assembly capacity and intermediate material. To this end, all stochastic influences are removed from the problem and it is formulated as a deterministic linear program. Then, the values of the dual variables corresponding with the resource constraints in an optimal solution are used as bid prices. Additional notation is summarized in Table 2.3.

In the considered problem, the future demand is uncertain. Therefore, the set of future incoming orders must be determined. In literature, next to the DLP, where random demand is simply replaced by expected values, another approach is to generate I $(i = 1, \ldots, I)$ demand scenarios with orders $\mathcal{D}_i$ according to the probability distributions. Next, the bid prices are computed for each $\mathcal{D}_i$. The used bid prices are determined by taking the mean of the bid prices computed with the demand scenarios $\mathcal{D}_i$. This approach is known as randomized linear programming (*RLP*); see, e.g., Talluri and van Ryzin (1999). We modify this approach and do not use a fixed number of demand scenarios. Instead, we iteratively update the bid prices using new generated demand scenarios until the mean over all computed bid prices converges. To avoid premature convergence if the first demand scenarios generate identical bid prices, we use a minimum number of demand scenarios. To the best of our knowledge, this approach has not yet been described in literature.

Assuming deterministic future demand, the problem at hand reduces to the problem of which orders to accept and how to schedule them. Quoting the due dates is trivial in this case because the optimal due dates correspond with the completion times of the orders; they are known after scheduling all orders. Therefore, we only need to schedule the start of assembly for each order. Let

$$prof_{dt} := contr_d - hc_{dt} - bc_{d(t+lead_d)} \tag{2.13}$$

define the marginal profit that an order d generates if its assembly begins in planning period t. In this definition, holding and backlog costs are subtracted from the contribution margin of the order. Next, we define the objective function for computing bid prices using the i-th

$\mathcal{D}_i$	Orders of demand scenario i
n_t	Number of accepted orders up to planning period t in the actual, real scenario
n_t^{bp}	Mean number of accepted orders up to planning period t in the bid price computation
bp_t^{cap}	Bid price for assembly capacity in t
bp_{mt}^{mat}	Bid price for intermediate material m in t
opp_{d,δ_d}^{due}	Opportunity costs for accepting d with due date δ_d
$opp_{d,t}^{res}$	Opportunity costs for starting assembly of d in t

Table 2.3: Notation for bid prices and opportunity costs

demand scenario as follows:

$$\sum_{d \in Y \cup \mathcal{D}_i} \sum_{t=arr_d+1}^{T-lead_d} prof_{dt} \cdot x_{dt}. \tag{2.14}$$

The restrictions (2.7)–(2.10) must apply to guarantee feasibility, where the constraints (2.7), (2.8), and (2.10) must consider all orders $d \in Y \cup \mathcal{D}_i$. The following additional set of inequalities ensures that more will not be produced than is demanded:

$$\sum_{t=arr_d+1}^{T-lead_d} x_{dt} \leq 1 \quad \text{for } d \in \mathcal{D}_i \tag{2.15}$$

In an optimal solution of the linear program (2.7)–(2.10) and (2.15) that maximizes (2.14), the values of the dual variables corresponding to the constraints (2.7) (i.e., bp_t^{cap}) give the bid prices for the assembly capacity in the respective planning period. Similarly, the values of the dual variables corresponding to the constraints (2.8) in an optimal solution (i.e., bp_{mt}^{mat}) give the bid prices for each type of intermediate material in each planning period.

2.6.2 Opportunity cost approximations

In this section, we show how to use the bid prices to approximate the opportunity costs of accepting an order with a given actual start of assembly and with a given due date, respectively.

First, we define the opportunity costs for accepting order d with *given actual start of assembly* in planning period t. To this end, we add up the bid prices of the single resources required by the order:

$$opp^{res}_{dt} := \sum_{m=1}^{M} use_{dm} \cdot bp^{mat}_{mt} + \sum_{i=0}^{lead_d - 1} bp^{cap}_{t+i}. \tag{2.16}$$

Note that these values are known if the assembly date is fixed. This approach is an analogue to the process in network revenue management in which more than one resource is also required to fulfill an order; see, e.g., Talluri and van Ryzin (1999).

Even with a given due date, it is not yet clear which resources will actually be used to produce an order because the actual assembly date is not fixed yet. To cope with this problem, we assume that the order can be started in the planning period with lowest opportunity costs. This approach is in line with literature and has been used in the context of FP-RM (Gallego et al. (2004), Petrick et al. (2012)) and revenue management in MTO manufacturing (Spengler et al. (2008)). To be specific, let

$$\mathcal{T}_{d\delta_d} := \{t \in \{arr_d + 1, \ldots, \delta_d - lead_d\} \mid t \text{ is feasible start of assembly}\}.$$

As the best feasible planning period for the start of the assembly can be chosen, we take the feasible planning period that minimizes the sum of the opportunity costs for the used resources and the resulting holding costs:

$$opp^{due}_{d\delta_d} := \min_{t' \in \mathcal{T}_{d\delta_d}} \{opp^{res}_{dt'} + hc_{dt'}\}. \tag{2.17}$$

2.6.3 Update bid prices

As demand is stochastic, updating the bid prices if new information becomes available can improve the results. We use the following methods to do so.

A simple approach to updating bid prices, which is also commonly used in literature (see e.g. Talluri and van Ryzin (1999)), is to re-compute them after a certain number of periods. In this work, we use the knowledge of the already accepted orders and their quoted due dates. Additionally, information on orders that are currently in the process of assembly and the resources required by them in future planning periods is available.

Furthermore, we adapt the bid prices in between successive updates. In the literature, there are some approaches that adapt bid prices without resolving (Adelman (2007), Klein (2007)). However, they require a rather high additional computational effort, while our approach is easy to implement and the required computation time is negligible. We collect information about how many orders are accepted in the virtual scenarios during the bid price computation. Let n_t^{bp} represent the mean number of accepted orders up to planning period t in the bid price computation, and let n_t denote the number of accepted orders up to planning period t in the actual, real scenario. At the end of planning period t, bp^{cap} and bp^{mat} are adjusted by multiplying them with n_t/n_t^{bp}, the ratio between the actual number of accepted orders and the expected number of accepted orders. Therefore, if fewer orders than expected were accepted, the bid prices decrease, and vice versa.

Additionally, at the beginning of each planning period, we set the bid price for the assembly capacity in the same period to 0 because no new arriving orders can begin assembly in this period.

2.7 Numerical study

In the following numerical study, we evaluate the performance of the previously described revenue management approach, which we refer to as ATO-RM. After presenting the test bed and the used benchmark algorithms, we compare the results of ATO-RM and the benchmark algorithms with an ex-post optimal solution. Next, a sensitivity analysis shows the influence of the varied parameters on the performance of the tested algorithms and on the structure of the results.

The simulation environment corresponds with the description of the approach in Algorithm 1. The algorithms were implemented in C++ using a Gurobi 5.5 solver on a 3.20 GHz Intel Core i7 machine with 32GB of RAM.

2.7.1 Test bed

First, we present the scenarios and algorithms used in the numerical study. The parameters for the scenarios are summarized in Table 2.4. For the sake of clarity of analysis, we model a basic situation.

Fixed parameters	
Number of planning periods in planning horizon T	60
Number of intermediate materials M	1
For all orders d:	
Coefficient for required int. material use_{dm}	1
Required number of consecutive periods of assembly capacity $lead_d$	7
Pref. due date $pref_d$ dependent on order type	$arr_d + (8/11/14)$
Holding cost rate c_d^H	1% of $contr_d$
Backlog cost rate c_d^B	10% of $contr_d$
Varied parameters	
Profit heterogeneity ($contr_d$ dependent on order type)	$\{(600/550/500),$ $(800/550/300),$ $(1000/550/100)\}$
Coefficient of variation cv	$\left\{\frac{1}{3}, \frac{2}{3}, 1\right\}$
Available assembly capacity a_t	$\{205, 175, 155, 135\}$
Int. mat. replenishment r_{mt} (in planning periods $t = 1, 16, 31, 46$)	$\{1600, 1360, 1180, 1040\}$
Demand arrival pattern (peaks)	{early,middle,no peak}

Table 2.4: Parameters used in the numerical study with three order types

Order arrivals

To be able to easily compare the considered problem in the current paper to models from the literature, in the SDP (2.5), the order arrival process is modeled by individual order arrival probabilities within each micro-period, which is the standard way to model the order arrival process in a SDP in existing revenue management literature (see e.g. Spengler et al. (2007), Petrick et al. (2012)). However, in practice, it is typically harder to determine suitable parameters for the order arrival probabilities in each micro-period than a probability distribution describing the number of orders arriving in a planning period. Therefore, in the numerical study we use this order arrival structure.

At the beginning of the simulation, for each planning period, the number of arriving orders of each of the three different existing order types is drawn according to a negative binomial distributed random variable $NB(\mu_t, cv)$ with mean μ_t and coefficient of variation cv. We choose this distribution because it is commonly used in literature (see, e.g., Ehrenberg (1959) or Agrawal and Smith (1996)) and allows for a high coefficient of variation. To reflect an adequate range of variability in the test cases, the coefficient of variation cv is varied. In accordance with the assumptions from Section 2.3, in the simulation process, the program learns the arriving orders only one after the other.

To observe how the algorithms perform in different demand arrival scenarios, scenarios with three different demand arrival patterns are generated:

- No peak: $\mu_t = 10 \;\text{ for }\; t = 1, \ldots, 46$
- Early peak: $\mu_t = \begin{cases} 13.5 & \text{for } t = 1, \ldots, 10 \\ 9 & \text{for } t = 11, \ldots, 46 \end{cases}$
- Middle peak: $\mu_t = \begin{cases} 9 & \text{for } t = 1, \ldots, 18 \text{ and } t = 30, \ldots, 46 \\ 13.5 & \text{for } t = 19, \ldots, 29 \end{cases}$

In the last 14 planning periods, no orders arrive ($\mu_t = 0$ for $t = 47, \ldots, 60$) such that all orders can be fulfilled within their preferred due dates.

Profits

To observe whether applying a revenue management approach is worthwhile even if the profits gained from different customer classes are nearly the same, the profit heterogeneity that reflects the different importance of customers is varied. High-value customers request shorter preferred lead times than low-value customers. This situation reflects the fact that customers generally must pay a premium to obtain shorter lead times, and short lead times are provided for important customers. That is, we consider three order types with preferred lead times of 8, 11, and 14 planning periods, respectively, and profits depending on the profit heterogeneity. The exact values are given in Table 2.4. The holding (backlog) cost rate for an order is 1% (10%) of its contribution margin.

Resources

Each order requires one unit of the only required intermediate material and seven consecutive planning periods of assembly capacity. The ratio of the congestion of resources is especially interesting because we model scarcity in both resources. To be able to adjust capacity and intermediate material congestion independently, we keep the expected number of order arrivals fixed over the entire time horizon and vary the number of available parallel machines a_t and the replenishment amount for the intermediate material r_{mt}, respectively. In this way, we generate scenarios with approximately 85%, 100%, 115%, and 130% congestion for each resource, respectively.

Overall, this leads to $3^3 \cdot 4^2 = 432$ scenario settings. For each scenario setting, 20 instances are randomly generated, resulting in 8640 instances.

Algorithms

For ATO-RM, the following settings are used to compute single resource bid prices. We begin by computing bid prices using ten demand scenarios generated according to the known probability distributions. In Section 2.6.1, we did not specify how we define convergence of the bid prices. In this numerical study, we check convergence by looking at the sets of bid prices generated by the five most recent integrated demand scenarios. We add the bid prices one by one and test if one of the mean bid prices changes by more than 1 absolute unit and 5% relative to the previous mean bid price. If this is not the case for all of the bid prices, we state that the bid prices have converged. Single resource bid prices are recomputed every 12 planning periods and adapted at the beginning of each planning period in which the bid prices have not been recomputed; cf. Section 2.6.3.

We compare ATO-RM with the following benchmark algorithms:

- First-come-first-served (FCFS): An order is accepted and is quoted its preferred due date if there are sufficient resources available to fulfill the order within the preferred due date, otherwise it is rejected. Accepted orders are scheduled as early as possible. Although the FCFS approach does not differentiate between different customers, it is still used in practice because it is a clear and easy to use strategy.

- Revenue management approaches that ignore one of the two resources in the bid price computation:

  - ONLYCAP ignores the intermediate material availability, i.e., inequality (2.8) is not included in the linear program when computing single resource bid prices and $bp_t^{mat} = 0$ for all t.

  - ONLYMAT ignores the assembly capacity restrictions, i.e., inequality (2.7) is not included in the linear program when computing single resource bid prices and $bp_t^{cap} = 0$ for all t.

  These algorithms are proxies for processes using existing MTO or MTS revenue management approaches.

We would prefer to compare the heuristic ATO-RM approach with an optimal policy. However, this comparison is impossible because computing an optimal policy is intractable for problems of the given size. Instead, we compare the results to an ex-post optimal solution: If all arriving orders are known, one can compute the maximum attainable profit by solving the deterministic linear program described in Section 2.6.1. Note that in practice this ex-post algorithm (say POSTOPT) cannot be implemented because the demand is unknown. POSTOPT gives an upper bound on the maximum attainable profit.

2.7.2 Simulation results

Next, we compare the results of ATO-RM and the benchmark algorithms with POSTOPT and explain the reasons for these results.

The main result is shown in Fig. 2.3, which illustrates the mean relative difference to the optimal ex-post profit over all 8640 tested instances for ATO-RM and the benchmark algorithms. We emphasize three observations:

- ATO-RM is close (3.96%) to the ex-post optimal solution indicating that the proposed approach works quite well. This result is even more impressive because the ex-post solution was created with full knowledge of all incoming demand, whereas ATO-RM

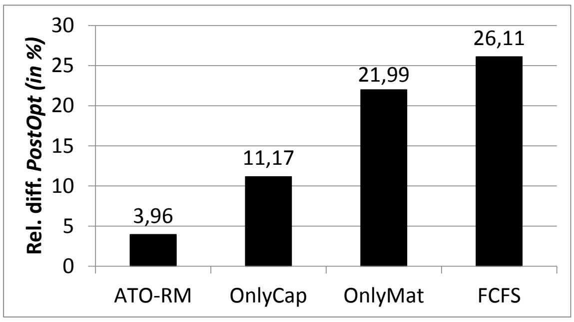


Figure 2.3: Mean relative differences to the optimal ex-post profit

used only information from the probability distributions, which means ATO-RM will produce a result even closer to a solution computed with an optimal policy, which also has only information from the probability distribution.

- ATO-RM clearly dominates the benchmark algorithms. It follows that order discrimination and taking into account both scarce resources are important to obtaining good results.
- ONLYCAP (11.17%) performs much better than ONLYMAT (21.99%), which indicates that in our test cases, considering the assembly capacity constraints appears to be more critical than considering the intermediate material constraints. Also if only scenarios are considered, in which the level of congestion is the same for both resources, ONLYCAP (5.65%) clearly outperforms ONLYMAT (16.77%). A more detailed analysis of the influence of congestion on the algorithms is given in Section 2.7.3.

Fig. 2.4 shows the distribution of the relative differences of the resulting profits to the ex-post optimal solution over all 8640 tested instances for ATO-RM and the benchmark algorithms. The above-described results do not only apply in the mean; ATO-RM also performs well in most of the tested scenarios, leading to an upper quartile of 5.28%. The benchmark algorithms are not only worse in terms of the mean, but there are also a significant amount of instances in which the resulting profits are located rather far away from the ex-post solution. The sharp bend in the curve for FCFS stems from the fact that, for low profit heterogeneity, FCFS performs quite decent because order selection is less important. 93.8% of the instances with low profit heterogeneity show a relative difference to POSTOPT of $9-15\%$. In Section 2.7.3, we illustrate the behavior of the algorithms in different scenarios in greater detail.

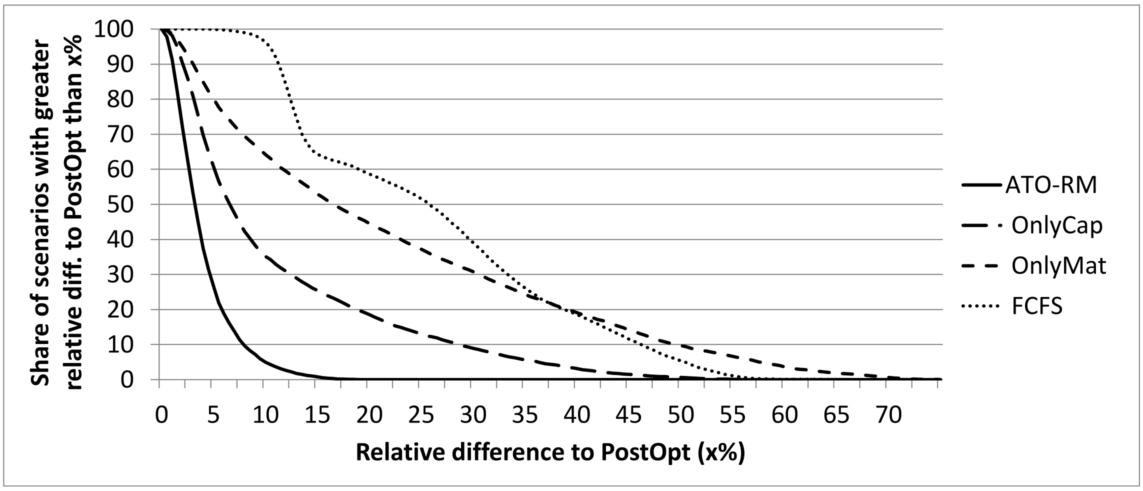
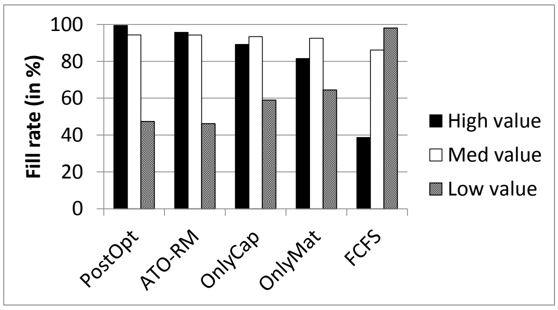
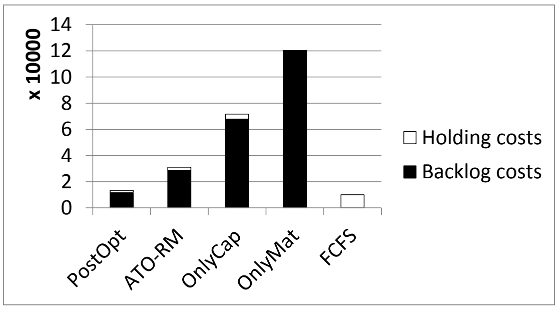


Figure 2.4: Distribution of the relative differences to the optimal ex-post profit over all instances



(a) Fill rate for the different order types



(b) Resulting holding and backlog costs

Figure 2.5: Structure of the solutions computed by the different algorithms

In the following, we provide additional insights to explain these results. The main decisions of the algorithms are which orders are accepted and when they are scheduled. Thus, we further examine the fill rate for the different order types and the resulting backlog and holding costs for each algorithm.

PostOpt accepts almost all high-value orders and only approximately half of the low-value orders as shown in Fig. 2.5a. The revenue management approaches exhibit similar behavior. Still, ATO-RM accepts a larger amount of orders from high-value customers than OnlyCap and OnlyMat. One reason for this observation may be that the bid prices computed by ATO-RM are generally higher than those of OnlyCap and OnlyMat because they ignore the opportunity costs for one of the two scarce resources. FCFS accepts a larger amount of low- than high-value orders because low value orders have longer preferred lead times. Therefore, in case of scarce resources, it is still feasible to accept a low-value order but not to accept a high-value order.

Fig. 2.5b shows the resulting backlog and holding costs. The POSTOPT solution results in relatively low backlog and holding costs, perhaps because this algorithm uses information on all future incoming demand and thus quotes no due dates later than the actual delivery date. ATO-RM generates lower costs than ONLYCAP and ONLYMAT. Again, one reason for this observation might be that these approaches underestimate the opportunity costs because they only take into account the opportunity costs of one of the two resources. This situation can lead to accepting orders with high backlog costs, whereas ATO-RM rejects these orders after comparing the resulting profits to the corresponding bid price. As was clear by definition, FCFS creates no backlog costs.

The bid prices for a given order computed by ONLYMAT at a certain point in time are essentially a decreasing step function of the quoted due date. The steps originate from the points in time at which new intermediate material becomes available. Therefore, late points in time are viewed as more appealing (cf. Section 2.6.2), which might be a reason why ONLYMAT generates much more backlogging than all other algorithms. Additionally, this situation might lead ONLYMAT to accept a larger amount of low-value orders due to low bid prices that do not take into account that not all orders can be assembled at that point in time because of the assembly capacity constraints. Thus, assembly capacity is not saved for high-value orders, which in turn must be rejected.

Computation time is not a critical factor in our tests. For each instance, ATO-RM takes 30 seconds on average for bid price computation and 1 second for feasibility checks, order selection, etc.

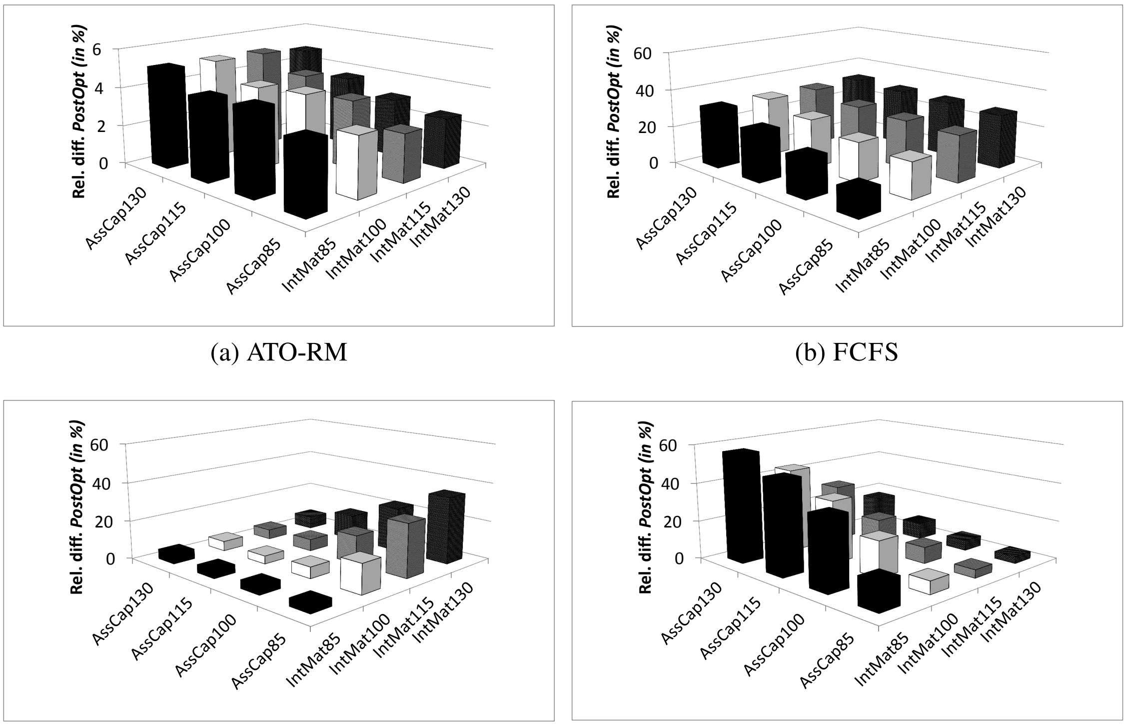
Additionally, we conducted numerical tests to show the robustness of our results. We increased the mean number of arriving orders, conducted tests with different resource consumptions of the orders, and we applied a slightly different cost structure. The numerical results from these tests support the findings presented in this paper. We also showed the statistical significance of the results by computing confidence intervals.

2.7.3 Sensitivity analysis

In the following, we examine the influence of the varied parameters on the performance of the algorithms as well as on the solution characteristics. All simulation results are

Instance		PostOpt	ATO-RM	OnlyCap	OnlyMat	FCFS
Overall		0(99.5,94.3,47.4,2.0)	4.0(95.7,94.3,46.2,4.7)	11.2(89.2,93.4,59.0,11.0)	22.0(81.5,92.5,64.5,19.0)	26.1(38.6,86.1,98.1,2.0)
ProfHet	Low	0(98.5,92.3,53.0,3.0)	3.0(90.0,90.6,59.9,4.0)	7.0(81.3,90.2,72.7,8.6)	16.6(72.9,89.3,76.5,15.8)	12.1(38.6,86.1,98.1,2.6)
	Med	0(99.9,94.8,46.3,1.9)	4.5(98.1,95.6,43.2,5.1)	12.8(91.4,94.1,57.0,12.0)	23.6(83.3,93.7,62.2,20.0)	26.8(38.6,86.1,98.1,2.0)
	High	0(100,95.7,42.9,1.2)	4.4(99.1,96.6,35.4,4.9)	13.6(94.8,95.8,47.3,12.1)	25.7(88.3,94.6,54.6,21.0)	39.4(38.6,86.1,98.1,1.4)
cv	1/3	0(99.9,96.3,45.7,1.8)	2.3(98.3,95.8,44.7,3.3)	11.1(90.5,94.4,59.1,10.7)	21.3(83.0,93.9,62.8,18.4)	26.6(37.3,87.1,98.2,2.1)
	2/3	0(99.6,94.9,48.3,1.9)	4.1(96.1,94.7,46.9,4.8)	11.3(89.5,93.8,60.1,11.0)	21.9(81.8,92.7,65.4,18.9)	25.8(39.1,87.6,98.3,2.0)
	1	0(99.0,91.8,48.2,2.3)	5.5(92.8,92.3,47.0,6.1)	11.2(87.4,92.0,57.9,11.1)	22.8(79.7,91.0,65.2,19.9)	25.9(39.5,83.7,97.8,2.0)
Congestion (cap%-mat%)	85-85	0(100,99.9,93.9,0.9)	3.5(98.4,99.4,94.7,4.1)	4.0(98.7,99.6,96.4,5.0)	14.1(97.6,99.3,97.9,14.8)	12.8(74.3,99.3,100,2.1)
	100-100	0(99.8,98.8,67.1,1.5)	4.4(96.3,98.3,66.0,4.9)	5.9(95.0,98.3,72.7,7.0)	18.9(88.7,96.0,78.8,17.7)	21.1(49.5,96.4,99.9,2.1)
	115-115	0(99.6,95.4,42.0,2.1)	4.3(95.2,95.3,39.8,4.9)	6.2(93.0,96.2,47.2,7.7)	16.6(86.2,93.3,53.4,15.1)	27.1(34.0,87.4,99.7,2.1)
	130-130	0(99.0,87.0,27.0,2.6)	4.8(94.1,87.6,25.6,5.4)	6.6(92.0,89.9,28.6,8.1)	17.5(82.6,87.3,35.6,15.1)	32.0(22.3,77.6,98.6,2.0)
	100-85	0(99.7,99.0,74.2,1.4)	4.3(96.3,98.5,72.6,4.5)	4.3(96.3,98.5,72.6,4.5)	36.7(83.9,94.6,92.8,34.1)	20.2(52.4,98.2,100,2.2)
	115-100	0(99.3,96.4,52.2,1.7)	4.2(95.6,96.6,46.6,4.4)	4.2(95.6,96.7,46.5,4.4)	32.8(75.1,91.4,74.4,28.1)	26.2(34.0,95.1,99.9,2.1)
	130-115	0(98.5,89.5,31.9,2.1)	5.1(94.3,90.2,27.5,5.1)	5.1(94.4,90.1,27.5,5.1)	29.5(70.7,88.2,51.4,23.9)	31.6(20.2,85.7,99.6,2.0)
	115-85	0(99.3,96.4,52.2,1.7)	4.2(95.6,96.7,46.5,4.4)	4.2(95.6,96.7,46.5,4.4)	48.4(61.7,91.4,90.3,41.4)	26.2(34.0,95.1,99.9,2.1)
	130-100	0(98.5,89.5,32.0,2.1)	5.1(94.2,90.2,27.5,5.1)	5.1(94.4,90.1,27.5,5.1)	43.6(55.4,87.1,70.2,35.7)	31.6(20.3,85.6,99.6,2.0)
	130-85	0(98.5,89.5,32.0,2.1)	5.1(94.3,90.1,27.5,5.1)	5.1(94.4,90.1,27.5,5.1)	56.4(39.2,86.5,88.1,45.2)	31.6(20.3,85.6,99.6,2.0)
	85-100	0(100,99.4,72.3,1.6)	3.1(97.0,98.6,74.1,4.1)	15.2(92.3,98.7,95.7,17.4)	6.7(95.6,98.5,80.8,8.3)	18.8(59.3,91.6,99.8,2.0)
	100-115	0(99.9,97.3,44.0,2.1)	3.5(95.9,96.6,45.8,4.7)	14.8(86.1,95.4,71.2,16.1)	8.5(92.9,96.3,54.7,10.0)	25.6(42.0,82.8,98.6,2.0)
	115-130	0(99.8,90.1,28.3,2.8)	3.7(95.1,90.9,29.6,5.3)	13.0(83.6,92.4,47.7,14.7)	8.7(90.6,90.9,36.5,9.9)	30.3(31.0,74.1,95.5,2.0)
	85-115	0(100,98.0,47.4,2.2)	2.5(97.3,96.9,49.5,4.1)	27.5(76.5,91.5,89.9,25.3)	4.3(95.7,96.8,54.5,6.2)	24.7(47.8,80.2,96.3,2.0)
	100-130	0(99.9,91.1,29.8,2.8)	3.0(95.9,91.2,31.7,4.9)	23.1(74.4,88.3,64.8,21.2)	5.5(93.3,91.4,36.2,7.3)	29.4(36.0,72.4,92.8,1.9)
	85-130	0(99.9,91.6,32.4,2.8)	2.6(96.2,91.0,34.0,4.3)	34.5(64.3,81.9,81.7,28.1)	3.7(94.8,91.3,35.9,5.3)	28.5(40.9,70.9,89.8,1.9)
Peaks	no peak	0(99.5,96.1,46.8,2.0)	3.8(95.8,96.0,45.4,4.5)	11.1(88.8,94.4,59.0,10.5)	22.1(81.0,92.8,65.2,18.6)	26.2(38.3,87.8,98.2,2.0)
	early	0(99.5,93.3,47.7,2.0)	4.2(95.3,92.9,47.7,4.9)	12.5(88.5,92.6,60.5,12.2)	21.1(83.2,92.1,63.9,19.0)	26.1(39.2,84.4,98.3,2.0)
	middle	0(99.4,93.5,47.8,2.0)	3.8(96.2,93.8,45.5,4.7)	10.0(90.3,93.1,57.5,10.1)	22.8(80.3,92.6,64.3,19.5)	26.0(38.3,86.1,97.8,2.0)

Table 2.5: Results from sensitivity analysis [DiffOpt(Fill rates, $hcbc\_share$)]



(a) ATO-RM

(b) FCFS

(c) ONLYCAP

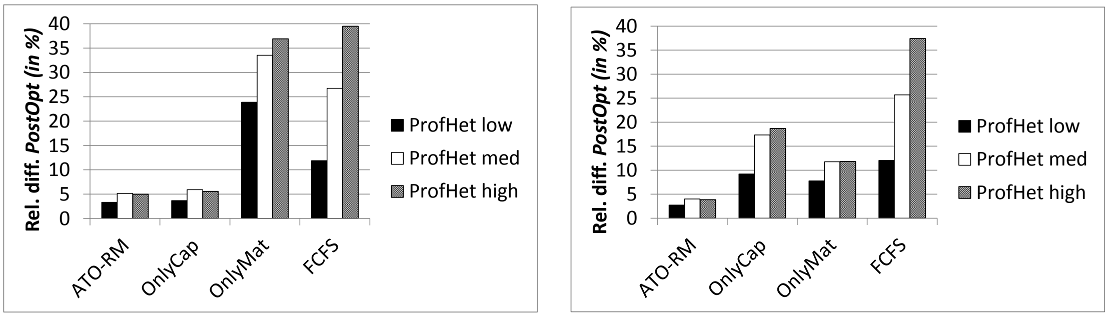
(d) ONLYMAT

Figure 2.6: Mean relative difference to the optimal ex-post profit under different resource congestions

summarized in Table 2.5, which shows the relative difference to POSTOPT and gives the fill rates for high-, medium-, and low-value orders in parentheses. Additionally, the $hcbc\_share := \frac{\text{holding costs + backlog costs}}{\text{holding costs + backlog costs + profit}}$ is presented, which gives the relationship between the costs and the contribution margin. A relative measure of the holding and backlog costs is necessary because a solution with more accepted orders (and thus more revenue) is expected to have also more holding/backlog costs. The results are generated by fixing one (or in case of different congestion, two) of the varied parameters and taking the mean over all scenarios with these fixed parameter values.

Impact of congestion

Fig. 2.6 depicts the influence of the different supply scenarios on the performance. Note the different scale in the diagram for ATO-RM. ATO-RM performs well under all circumstances, but its performance worsens with increasing scarcity of assembly capacity. This situation corresponds with the observation that consideration of assembly capacity appears to be



(a) Scenarios, in which assembly capacity is at least as scarce as intermediate material

(b) Scenarios, in which intermediate material is at least as scarce as assembly capacity

Figure 2.7: Mean relative difference to the optimal ex-post profit under different levels of profit heterogeneity

more important than consideration of intermediate material (cf. Section 2.7.2). Note that ATO-RM performs better than ONLYMAT even if intermediate material is the clear bottleneck resource and as good as ONLYCAP if assembly capacity is the clear bottleneck resource, which is in line with the results reported in Section 2.7.2 . As expected, if the bottleneck is not taken into account in the benchmark algorithm, the respective algorithm performs poorly. FCFS performs worse the more congested the system becomes.

The reason why ATO-RM performs worse if assembly capacity is the bottleneck resource might be because intermediate material is required only at the beginning of assembly, whereas assembly capacity is needed over several planning periods, and it is thus harder to manage. Additionally, assembly capacity is not storable, and therefore unused capacity is ultimately lost. For ONLYCAP and ONLYMAT, it can be observed that in the cases that ignore the bottleneck resource, the algorithms do not differentiate between order types anymore, perhaps because they do not see a bottleneck and thus do not reject orders due to the bid prices. This also may be an explanation why the backlog costs drastically increase in these cases. If the congestion is still low, additional orders can be accepted, and thus FCFS also accepts a large amount of high-value orders, which can be a reason for its performance curve.

Due to the completely different results of ONLYCAP and ONLYMAT under different resource congestions, in the remaining section, we divide the analysis by the dominating bottleneck of the scenarios.

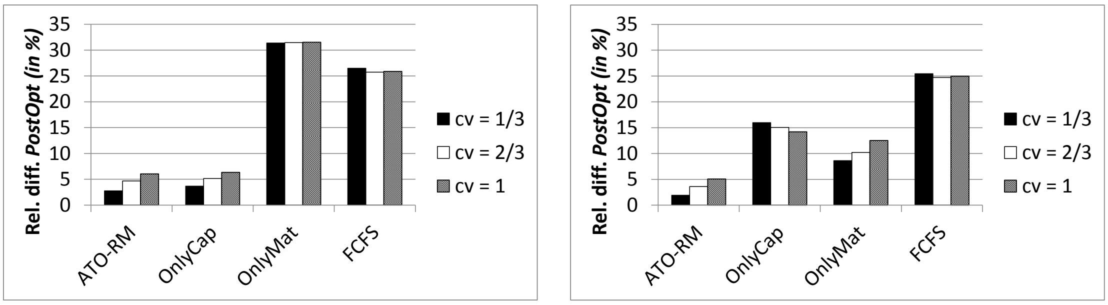
Impact of profit heterogeneity

Fig. 2.7 shows the influence of profit heterogeneity on the performance of the tested algorithms. It can be observed that using ATO-RM is worthwhile even if the profit heterogeneity is low. Increasing profit heterogeneity in our tests leads to worse results for FCFS, whereas ATO-RM is influenced to a lesser degree. The influence of increasing profit heterogeneity on ONLYCAP and ONLYMAT is higher in those scenarios, in which the considered resource is not the bottleneck resource than in the scenarios, in which the modeled resource is dominant.

With the increasing difference of the profits from orders of different customer classes it becomes increasingly more important to select which orders are accepted. For ATO-RM, it is easier to differentiate between the order types with increasing profit heterogeneity, which means that even though the bid prices might be not completely correct, they can lead to the correct decisions. Therefore, the fill rates of high- and medium-value orders are notably high. However, the overall number of accepted orders is relatively low due to the rather low acceptance rate for low-value orders. The same holds for ONLYCAP and ONLYMAT in those scenarios, in which the considered resource is the bottleneck resource. FCFS suffers most from increasing profit heterogeneity because the fill rates for all customer types remain the same regardless of the profit heterogeneity. This holds also true for ONLYCAP and ONLYMAT in the scenarios, in which the considered resource is of subordinate importance because in these scenarios, bid prices are rather low and thus order selection is less strict. Due to the high backlog costs, FCFS performs better for low profit heterogeneity than ONLYMAT in those scenarios, in which assembly capacity is the bottleneck resource.

Impact of coefficient of variation

Fig. 2.8 shows the influence of the coefficient of variation of the arriving demand on the performance of the algorithms. In all of the scenarios, ATO-RM appears to suffer from an increasing coefficient of variation, whereas FCFS is relatively indifferent in this case. However, even in the worst tested case, ATO-RM still clearly dominates the benchmark algorithms. ONLYCAP and ONLYMAT seem to suffer from an increasing coefficient of variation in those scenario, in which they model the bottleneck resource, and they are indifferent in the other scenarios.



(a) Scenarios, in which assembly capacity is at least as scarce as intermediate material

(b) Scenarios, in which intermediate material is at least as scarce as assembly capacity

Figure 2.8: Mean relative difference to the optimal ex-post profit under different coefficients of variation

The reason for the decreasing performance of ATO-RM may be that the bid prices computed by ATO-RM are generated using many different demand situations. In the case of a high variance, these situations differ to a greater extent. Therefore, a scenario in which the computed bid prices do not fit the actual demand situation might occur more often, which might be the reason for increasing $hcbc\_share$ and a decreasing fill rate for high-value orders, leading to worse results of ATO-RM. The same holds true for OnlyCap and OnlyMat in those scenarios, in which they consider the bottleneck resource and thus compute reasonable bid prices. In the other scenarios, OnlyCap and OnlyMat do not use the correct bid prices in the first place because they ignore one of the limiting resources. It is possible that this is why a higher variance does not significantly influence their performance in these cases. FCFS makes no use of forecasts at all. As the mean demand remains the same, the solution quality is also nearly the same. For all tested algorithms except FCFS, the fill rates for high- and medium-value orders decrease with higher variance, which may be related to the arriving demand being less uniform over time.

Impact of demand arrival patterns

From the results, no significant influence of different demand arrival patterns on the tested algorithms can be observed.

2.8 Conclusions and further research

In the present paper, we model the decisions of an ATO manufacturer. In addition to the decisions on accepting orders, we also include scheduling and due date quoting decisions. The intermediate material and assembly capacity are explicitly modeled in this work. We present a novel revenue management approach that uses bid prices to make the decisions.

As shown from the numerical study, ATO-RM works quite well. On average, the profit obtained is close to that of the ex-post solution, which contains full knowledge of all incoming demand. The proposed approach clearly outperforms the other tested benchmark algorithms. In our tests, the capacity constraints appear more important than the intermediate material constraints. This observation is illustrated by the better performance of ONLYCAP in comparison with ONLYMAT and by the decreasing performance of ATO-RM with a higher scarcity of assembly capacity. Computation time is no issue in the tested instances.

The sensitivity analysis shows that increasing the variance and scarcity of the assembly capacity lead to decreasing quality of ATO-RM. However, even in these situations, ATO-RM still performs quite well. The proposed approach outperforms the benchmark algorithms in all of the tested settings.

The developed approach can be generalized to fit other settings.

For example, if customers are only informed about acceptance or rejection of their order, backlog costs depend only on the difference of the preferred and the actual due date. Therefore, only the order acceptance decision must be made. The due date quoting is trivial for every accepted order. In the scheduling decision, we need to consider that backlog costs occur if an order is finished after the preferred due date. Therefore, backlog costs must be included in the scheduling decision in a similar way as holding costs are considered.

Another example is that orders must be produced as a whole. In this case, the variables for the assembly decisions in the LP used for feasibility checks and scheduling must be restricted to integer values. Note that we cannot assume this in the bid price computation as otherwise no shadow prices can be derived.

Also, if the number of products within an order varies, this can be easily incorporated by introducing a new parameter.

In the last example, we assume that customers adapt to the quoted due date and thus holding costs depend on the quoted due date instead of the preferred due date. In this case, we can just change the definition of holding costs. However, the modeled production environment is still rather basic. Order-dependent set-up costs are not considered. We assume one-stage deterministic production and deterministic replenishments of the intermediate materials. Relaxing these assumptions offers a good opportunity for further research. It also would be of interest to analyze how the algorithms perform under more complex material structures or in a system with additional stochastic influences.

3 Revenue management in a multi-stage make-to-order production system

Co-authors:

Moritz Fleischmann

Chair of Logistics and Supply Chain Management, Business School, University of Mannheim, Germany

Lars Mönch

Chair of Enterprise-wide Software Systems, Department of Mathematics and Computer Science, University of Hagen, Germany

Raik Stolletz

Chair of Production Management, Business School, University of Mannheim, Germany

Working paper.

Abstract:

In the present paper, we consider the demand management decisions of a manufacturer facing stochastic demand. While in the previous literature, either the order release decisions are fixed upon arrival or a single-stage production environment is assumed, we make online order acceptance and order release decisions in a multi-stage production system. After describing the problem formally as a stochastic dynamic program, we develop a bid-price-based revenue management approach in which the order acceptance and order release decisions are made based on previously computed bid prices and feasibility according to a linear program describing the multi-stage production system. A numerical study shows the good performance of the approach compared to an ex-post optimal solution in various supply and demand settings as well as the benefits relative to existing models, which either consider aggregated single-stage production or fix order release decisions upon order arrival.

3.1 Introduction

Revenue management has emerged mainly in the service industry, and most of the literature on the subject also involves this industry; see Talluri and Van Ryzin (2004) for an overview. Nevertheless, revenue management ideas have also been applied in make-to-stock (MTS) production systems (Quante et al., 2009; Meyr, 2009), make-to-order (MTO) production systems (e.g., Easton and Moodie, 1999; Barut and Sridharan, 2005; Kuhn and Defregger, 2005; Spengler et al., 2007; Watanapa and Techanitisawad, 2005; Spengler et al., 2008; Modarres and Sharifyazdi, 2009; Hintsches et al., 2010; Li et al., 2012; Volling et al., 2012; Chevalier et al., 2015), and assemble-to-order (ATO) production systems (Harris and Pinder, 1995; Gao et al., 2012; Guhlich et al., 2015b).

In the present paper, we consider an MTO manufacturer using a multi-stage production system. Short-term capacity adjustments are impossible during the considered planning horizon, and the available capacities are tightly planned. Demand is uncertain, fluctuating, and exceeds available capacities during peak phases. Therefore, the scarce resources that are available must be allocated to incoming demand; that is, order acceptance and order release decisions must be made. These decisions are of crucial importance if orders are heterogeneous, i.e., when they differ in capacity requirements and contribution margins.

This is the case, e.g., in the steel industry (Hintsches et al., 2010), where different products are produced to order in a multi-stage no-wait flow shop, and capacity expansion is very costly. Orders differ in their capacity consumption and contribution margins. It is possible to reject orders from customers without long-term contracts because each of these customers has a relatively low overall economic impact.

Another possible field of application is in semiconductor backend facilities in which chips are customized in a multi-stage production system, based on customer-specific requirements (Brown et al., 2000). In this production environment, capacities are also fixed and tightly planned due to high capital investment costs. Again, demand is fluctuating and uncertain.

While many revenue management approaches only consider order acceptance decisions (e.g., Talluri and van Ryzin, 1999; Spengler et al., 2007; Hintsches et al., 2010), there is typically more than one possible release date for an incoming order in the described production environments. Keeping more flexibility in the production schedule allows the company to accept and produce more orders.

Guhlich et al. (2015b) address this issue and consider online order acceptance and order release decisions in a single-stage production system. However, the production system is very simplistic and it is unclear how to handle a richer multi-stage system. This is the issue that we address in this paper. We show that applying the single-stage approach to an aggregation of a multi-stage system leads to poor results in general. Thus multi-stage systems require a dedicated approach.

In the present paper, we consider the demand management decisions of an MTO manufacturer using a multi-stage production system under stochastic demand. Incoming orders are accepted (or rejected) immediately upon arrival, while order release decisions are made periodically for the pool of accepted orders. The goal is to maximize the marginal profit generated by the accepted orders, while taking into account holding and backlog costs. In the existing literature, either a single-stage production system is considered, or order release decisions are fixed upon order arrival.

The main contributions of the present paper are as follows:

- We introduce a novel revenue management problem, considering online order acceptance and online order release planning in a multi-stage production system, and we model it as a stochastic dynamic program (SDP).
- We develop a heuristic revenue management approach based on bid prices to solve instances of realistic size.
- A numerical study shows the good performance of our approach in different production environments in comparison to an ex-post optimal solution. The study suggests that aggregating the considered multi-stage production systems into single-stage systems and then applying the approach described in Guhlich et al. (2015b) leads to unsatisfactory results. In addition, making *online* order release decisions can lead to significantly greater profits.

The paper is organized as follows. In the next section, an overview of the related literature is given. The problem is formulated in Section 3.3. In Section 3.4, we model the problem as an SDP. We develop a bid-price-based revenue management approach in Section 3.5. A numerical study evaluates the performance of the presented heuristic in Section 3.6. Section 3.7 concludes the paper.

3.2 Literature review

In this section, we give a brief overview of bid-price approaches in the revenue management literature. We consider applications in the service and manufacturing industries. In particular, we concentrate on the supported decisions and the production systems considered.

Network revenue management is concerned with managing multiple resources, such as different flight legs (see e.g., Talluri and Van Ryzin, 1998). There is no flexibility to assign customers to different resources (flight legs) after the purchase. Therefore, it is impossible to perform online scheduling in the previous literature, as is done in the present paper.

Revenue management of flexible products (Gallego and Phillips, 2004; Gallego et al., 2004; Petrick et al., 2010, 2012; Gönsch et al., 2014) allows the company to delay the assignment of customers to specific products (e.g., a specific flight). These decisions are made only after all requests for the flexible products have arrived, which enables the company to use capacities more efficiently. In contrast, in the problem considered in the present paper, service provision and the arrival of orders cannot be separated because these events overlap. Due to this different time structure, these approaches are not applicable in an online order acceptance and order release planning model. An extensive discussion of this topic can be found in Guhlich et al. (2015b).

In revenue management in MTO manufacturing, bid-price approaches have been proposed in the context of steel production. Spengler et al. (2007) consider a multi-stage production model to compute bid prices via a multi-dimensional knapsack formulation. However, only one production start date is available for each arriving order. Therefore, no scheduling decisions are made. In addition, orders released in different planning periods do not share any common resources, which is a major difference to the model used in the present paper. Spengler et al. (2008) propose an extension to this approach that allows for shifting demand between different planning periods. However, the scheduling decisions must be made immediately upon arrival of the order, whereas the online scheduling approach in the present paper can change the release date of an order until its production is actually started. Hintsches et al. (2010) consider long planning periods so that a planning period is long enough that an order finishes production within the same planning period in which it was started. Again, there is only one possible production start date per order.

Guhlich et al. (2015b) develop a bid-price-based revenue management approach for ATO manufacturing. In their setting, production capacity and intermediate materials are scarce resources. However, in contrast to the present paper, they consider a single-stage production model, which makes their linear program describing the production system not applicable in the current paper. Consequently, also the aggregation of bid prices to opportunity costs for accepting or producing an order must be altered in the current paper. In addition, the order processing is different. Guhlich et al. (2015b) quote a firm due date for each accepted order, which determines the backlog costs, and the holding costs depend on the actual finish date of the order. In the current paper, holding and backlog costs depend only on the finish date of the order, which implies that the order acceptance and order release decisions must be adapted.

To sum up, no approach from the existing literature is capable of making online order acceptance and online order release decisions in a multi-stage production environment. Either the order release decisions are fixed upon order arrival, or a single-stage production environment is assumed.

3.3 Problem formulation

In this section, we state the basic assumptions of the considered planning problem. We describe the production system, the production capacity, the orders, the decisions, and the cost structure. Table 3.1 summarizes the notation.

The planning horizon is divided into *planning periods*, $t = 1, \dots, T$. Each planning period t is further divided into *micro-periods* (t, s) $(s = 1, \dots, S)$ such that in each of these micro-periods, one order arrives at most. Let $d = 1, \dots, D$ denote the incoming orders. We make the following assumptions:

Assumption 3.3.1 (Production system) *We consider a no-wait, flexible flow shop production system consisting of machine groups $m = 1, \dots, M$, with a_{mt} identical parallel machines available for machine group m in planning period t.*

In a no-wait production system, orders are processed without waiting after their release. This requirement stems either from technical factors, as in the steel industry (Wismer, 1972; Tang et al., 2001), or from the fact that no intermediate storage is available. Hall and Sriskandarajah (1996) give an overview of industries using this type of production system.

Assumption 3.3.2 (Production capacity) *Each order can only be processed by one machine at a time. The production of an order d requires cap_{dim} time units of capacity*

Indices:	
$t = 1, \ldots, T$	Planning periods
$s = 1, \ldots, S$	Micro-period s in a planning period
$d = 1, \ldots, D$	Orders
$m = 1, \ldots, M$	Machine groups
Parameters:	
arr_d	Arrival period of order d
$pref_d$	Due date for order d
$lead_d$	Lead time of order d (measured in planning periods)
cap_{dim}	Capacity (relative to length of planning period) that order d requires in the i–th period after release at machine group m
a_{mt}	Number of parallel machines of machine group m available in planning period t
$cap_t(X)$	Required capacity for orders X starting production in planning period t
$contr_d$	Contribution margin of order d
c_d^H, c_d^B	Holding/Backlog costs (per planning period) for order d
hc_{dt}, bc_{dt}	Holding/Backlog costs for order d if released in t
$prof_{dt}$	Marginal profit that order d generates if released in planning period t
p_{dts}	Probability that order d arrives in micro-period (t, s)
p_{0ts}	Probability that no order arrives in micro-period (t, s)
State variables:	
A'	Remaining available production capacity of all machine groups in all future planning periods
Y	Set of orders that are accepted but not released yet
$\mathcal{Z}$	Set of all feasible states
(A', Y)	State of the system given by remaining available capacity A' and orders Y that have been accepted but not yet released
$V_{ts}(A', Y)$	Maximum expected marginal profit-to-go in state (A', Y) in planning period t if $s = 0$ and micro-period (t, s) otherwise
Decision variables:	
x_{dt}	1 if order d is released in planning period t; 0 otherwise
X	Set of released orders in the current planning period

Table 3.1: Notation

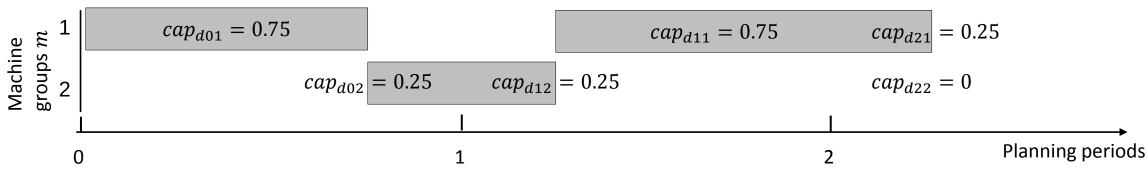


Figure 3.1: Example of a Gantt chart for an order d with lead time $lead_d = 3$

(relative to the length of a planning period, i.e., $0 \leq cap_{dim} \leq 1$*) in the* i*-th planning period after starting production at machine group* m*. The lead time,* $lead_d$*, is measured by the number of required planning periods. See Figure 3.1 for an example. Setup times are not considered, and processing times are deterministic.*

Let A' *denote the matrix of the remaining free production capacities in the remaining planning periods at all machine groups. For a set of orders* X *released in planning period* t*,* $cap_t(X)$ *is defined as the matrix of required production capacities (in all future planning periods at all machine groups). Its* (i, m)*-th entry describes the total capacity consumption at machine group* m *in period* $t + i$ *and is defined as* $\sum_{d \in X} cap_{dim}$*.*

Assumption 3.3.3 (Orders) *Orders arrive according to a known probability distribution. The probability that order* d *arrives in micro-period* (t, s) *is given by* p_{dts}*. The probability that no order arrives in micro-period* (t, s) *is given by* p_{0ts}*. Each order* d *has the following characteristics:*

- *Planning period that contains the micro-period in which the order arrived (*arr_d*)*
- *Due date (*$pref_d$*)*
- *Capacity requirement at machine group* m *in the* i*–th period after release (*cap_{dim}*)*
- *Lead time (*$lead_d$*)*
- *Holding and backlog cost rates (*c_d^H*,* c_d^B*)*
- *Contribution margin (*$contr_d$*).*

Assumption 3.3.4 (Decisions) *The company makes the following decisions:*

1. Order acceptance: *In each micro-period, the incoming order is immediately accepted or rejected.*
2. Order release: *At the beginning of each planning period, the set of orders to be released,* $X \subseteq Y$*, is selected.*

Here, Y denotes the set of orders that are currently accepted but not yet released. Orders can only be fully accepted and fully released to the shop floor. Orders can be released at the earliest in the first planning period after their arrival.

Assumption 3.3.5 (Cost structure) *If an accepted order is not fulfilled by the due date, backlog costs are incurred for each planning period beyond this date. Holding costs are incurred for orders finishing before the due date.*

If production for order d starts in planning period t, holding costs hc_{dt} and backlog costs bc_{dt} for this order are given by:

$$hc_{dt} := \max\{0, pref_d - (t + lead_d)\} \cdot c_d^H \tag{3.1}$$

$$bc_{dt} := \max\{0, (t + lead_d) - pref_d\} \cdot c_d^B. \tag{3.2}$$

If the production of an order d starts in planning period t, the following marginal profit is earned:

$$prof_{dt} := contr_d - hc_{dt} - bc_{dt}. \tag{3.3}$$

3.4 Stochastic dynamic program

In this section, we formalize the aforementioned problem by formulating it as an SDP. Specifically, we describe the state space, the decisions and transitions, the Bellman equation, and the feasibility check.

State space

The state of the system is given by (A', Y), the remaining available production capacity A' at each machine group in the upcoming planning periods and the set of orders Y that are accepted but not yet released.

Let $\mathcal{Z}$ define the set of all feasible states. A state is feasible if all accepted orders can be produced within the planning horizon. This can be checked using linear constraints, as defined at the end of this section.

Decisions and transitions

In each *micro*-period (t, s), the decision of whether to accept the arriving order is made. If the order is rejected (or no order arrives), the state remains unchanged. Otherwise, the accepted order d is added to the set of orders that are accepted but not yet released, denoted by $Y \cup d$, and the contribution margin of the order is earned.

At the beginning of each *planning* period t, the set of orders released $X \subseteq Y$ is determined. These orders are removed from Y (denoted by $Y \setminus X$), and the available capacities are updated (denoted by $A' - cap_t(X)$). Additionally, holding and backlog costs are determined for newly released orders.

Profits

Let $V_{ts}(A', Y)$ be the maximum expected marginal profit-to-go in state (A', Y) at the beginning of planning period t if $s = 0$ and in micro-period (t, s) for $s = 1, \dots, S$, otherwise. Additionally, we define $V_{t,S+1}(A', Y) := V_{t+1,0}(A', Y)$. Then, V_{ts} can be computed recursively via the Bellman equation for $(A', Y) \in \mathcal{Z}$ as follows:

$$V_{ts}(A', Y) =$$

$$\begin{cases} \sum\limits_{d} p_{dts} \cdot \max\{\overbrace{V_{t,s+1}(A', Y \cup d) + contr_d}^{accept\ d}, \overbrace{V_{t,s+1}(A', Y)}^{reject\ d}\} + \overbrace{p_{0ts} \cdot V_{t,s+1}(A', Y)}^{no\ order} & \text{if } 1 \leq s \leq S \\ \underbrace{\max\limits_{X \subseteq Y}\{V_{t,s+1}(A' - cap_t(X), Y \setminus X) - \sum_{d \in X}(hc_{dt} + bc_{dt})\}}_{order\ release\ decision} & \text{if } s = 0 \end{cases} \tag{3.4}$$

with boundary conditions $V_{T,S+1}(A', Y) = 0$ for $(A', Y) \in \mathcal{Z}$ and $V_{ts}(A', Y) = -\infty$ for $(A', Y) \notin \mathcal{Z}$.

We can use this SDP (3.4) to compare the considered problem with other models from the literature. The main difference from models in the literature on MTO-RM (Spengler et al., 2007; Hintsches et al., 2010) and FP-RM (Gallego et al., 2004; Petrick et al., 2010, 2012) is that in the SDP (3.4), there are two different time levels at which decisions are made. Order acceptance decisions are made immediately after order arrival, while order release decisions are only made at the beginning of each planning period. Compared to Guhlich et al. (2015b), in the present paper, multiple machine groups are considered, while Guhlich et al. (2015b) assume a single-stage production system. However, in the present paper, no intermediate materials are considered, and no due dates are quoted upon order arrival. Additionally, orders can only be fully released.

Feasibility check

To check the feasibility of a state, that is to check whether $(A', Y) \in \mathcal{Z}$ in planning period t^{fix}, we use a set of linear constraints. To this end, we define decision variables x_{dt} for each order d and planning period t with $\max(arr_d, t^{fix}) + 1 \leq t \leq T - lead_d$, which model the decision if order d is released in planning period t. The following constraints must be satisfied: [1]

$$\sum_{d \in Y} \sum_{i=0}^{lead_d - 1} cap_{dim} \cdot x_{d,t-i} \leq a'_{mt} \qquad \text{for } m = 1, \ldots, M; \ t = t^{fix} + 1, \ldots, T \tag{3.5}$$

$$\sum_{t=arr_d+1}^{T-lead_d} x_{dt} = 1 \qquad \text{for } d \in Y \tag{3.6}$$

$$x_{dt} \in \{0, 1\} \qquad \text{for } d \in Y; \ t = t^{fix} + 1, \ldots, T. \tag{3.7}$$

Constraint (3.5) describes the capacity consumption. The used capacity, which is composed of the different orders that may require machine group m in different planning periods after their start, must not exceed the available capacity. Constraint (3.6) guarantees that orders that have been accepted in previous planning periods will be finished within the planning horizon. Constraint (3.7) ensures that orders can only be fully released.

Constraints (3.5)-(3.7) are also the basis for the revenue management approach presented in the next section.

3.5 Revenue management approach

Because of the high dimensionality of the state space, it is computationally intractable to solve the SDP described in the previous section for instances of realistic size (cf. Petrick et al., 2012). Therefore, in this section, we develop a bid-price-based revenue management approach to solve the problem heuristically.

[1] We follow the convention that undefined variables are equal to 0.

The general approach is as follows:

1. At the beginning of the planning horizon, bid prices are computed using randomized linear programming (RLP) (Talluri and van Ryzin, 1999). Bid prices are used as a measure for the opportunity costs of consuming resources.
2. Upon arrival, we decide whether to accept an incoming order based on the previously computed bid prices and the feasibility according to the modeled production system. An order is accepted if there is a feasible release date for which the incremental profit exceeds the estimated opportunity costs of the required resources. The feasibility check is performed using constraints (3.5) - (3.7).
3. At the beginning of each planning period, order release decisions are made for the current planning period, considering all orders that are accepted but not yet released. To this end, a suitable objective function (using bid prices) is defined for constraints (3.5) - (3.7).

In the following, we describe in detail how to derive bid prices and how to use them to make order acceptance and order release decisions.

3.5.1 Bid price computation

To derive bid prices, the problem is modeled as a profit-maximizing non-integer linear program where future demand is assumed to be known. As known from the literature (Simpson, 1989; Williamson, 1992; Talluri and Van Ryzin, 2004), the derived primal optimal decisions can be discarded, and the shadow prices of the variables corresponding to the resource constraints are used as bid prices. They are a measure for the opportunity costs of using the corresponding resources. In the RLP, several demand scenarios $\mathcal{D}_j$ are generated, and the mean over the resulting bid prices is taken (Talluri and van Ryzin, 1999). We update the bid prices iteratively using new demand scenarios until the average over all computed bid prices converges, as described in Guhlich et al. (2015b). Hence, the number of scenarios is not fixed beforehand.

For scenario j, the objective function to be maximized is defined as follows:

$$\sum_{d \in Y \cup \mathcal{D}_j} \sum_{t=arr_d+1}^{T-lead_d} prof_{dt} \cdot x_{dt}. \tag{3.8}$$

In addition to Constraints (3.5) and (3.6), for each scenario j, the following inequalities must hold to guarantee that not more is produced than is demanded:

$$\sum_{t=arr_d+1}^{T-lead_d} x_{dt} \leq 1 \qquad \text{for } d \in \mathcal{D}_j. \tag{3.9}$$

We no longer force the decision variables x_{dt} to be integer because doing so would hinder shadow prices from being easily derived from the linear program. Relaxing the integrality constraints is also common practice in the classical revenue management approaches in the airline industry (cf. Talluri and Van Ryzin (2004)). Constraints (3.7) are therefore relaxed to $x_{dt} \geq 0$ for all orders $d \in Y \cup \mathcal{D}_j$, which allows for releasing fractions of an order in different planning periods.

The dual variables corresponding to the capacity constraints (3.5) in an optimal solution are then used as scenario bid prices $bp_j(m,t)$ for using machine group m in planning period t in scenario j. Production capacity that is unused in the current planning period is lost so that bid prices concerning these resources can be set to zero. The overall bid prices $bp(m,t)$ are then determined as the average over all scenario bid prices $bp_j(m,t)$.

3.5.2 Order acceptance decision

Similar as in Petrick et al. (2012), we define the opportunity costs of accepting order d and releasing it in planning period t as the sum of the bid prices of all required resources:

$$opp^{res}(d,t) := \sum_{m=1}^{M} \sum_{i=0}^{lead_d-1} cap_{dim} \cdot bp(m, t+i). \tag{3.10}$$

An order d is accepted if there exists a planning period t in which it is feasible to release d and

$$prof_{dt} \geq opp^{res}(d,t). \tag{3.11}$$

At this point in time, the order release date is still subject to change. Note that this approach underestimates the resulting backlog costs if accepting order d increases the backlog costs of one of the already accepted orders Y.

3.5.3 Order release decision

At the beginning of each planning period t, we must decide which of the accepted orders to release onto the shop floor. Similar as in Guhlich et al. (2015b), we use the opportunity costs to take into account future arriving demand. In particular, we maximize

$$\sum_{d \in Y} \sum_{t=t^{fix}}^{T-lead_d} (prof_{dt} - opp^{res}(d,t)) \cdot x_{dt} \tag{3.12}$$

with respect to constraints (3.5) - (3.7) to obtain a preliminary schedule. Only the order release decisions in the current planning period t^{fix} are implemented, while order release decisions for future periods are discarded. At the end of each planning period, the remaining available production capacity A' is updated.

3.6 Numerical study

In this section, we numerically investigate the performance of the revenue management approach presented in the previous section. First, we describe the experimental design. Next, we show that the proposed revenue management approach works well. Finally, the impact of aggregating the production systems to single-stage systems and the impact of making order release decisions online is examined.

3.6.1 Experimental design

In the following, we describe the parameter settings, production systems, and algorithms used in the experiments.

Parameter settings

In all simulations, we consider a planning horizon of $T = 40$ planning periods. Three different order types are available that all require the same resources but differ in their contribution margins. The due date depends on the order type. Order types that are more valuable arrive with an earlier due date because the willingness to pay is higher for customers with urgent orders. Therefore, high value orders can be delayed by only one planning period before incurring backlog costs. The due date for medium value orders is

Factor	Level	Count
Contribution margin $contr_d$ dependent on order type	$\{(200/150/100),$ $(300/200/100)$ $(500/300/100)\}$	3
Scarcity of capacity	$\{100\%, 110\%, 120\%\}$	3
Coefficient of variation CV	$\{0.5, 0.75\}$	2

Table 3.2: Parameters used in the numerical study

two planning periods after the earliest possible finish date, and low value orders can be delayed by four planning periods without incurring backlog costs. We obtain:

$$pref_d := \begin{cases} arr_d + lead_d + 1 & \text{for high value orders} \\ arr_d + lead_d + 2 & \text{for medium value orders} \\ arr_d + lead_d + 4 & \text{for low value orders.} \end{cases} \tag{3.13}$$

Holding and backlog cost rates are 3% and 5% of the contribution margin, respectively.

The performance of the algorithm depends on the considered demand scenarios. In particular, we expect that the profit heterogeneity of the orders, the scarcity level of capacity, defined as the ratio between demand and available capacity, and the coefficient of variation of the probability distribution for the incoming demand have a strong influence. Therefore, we vary these parameters as summarized in Table 3.2.

The profit heterogeneity in a demand scenario depends on the contribution margins for the high/medium/low-value orders. In the scenarios with low profit heterogeneity, the contribution margin of a high-value order is twice the contribution margin of a low-value order, while in a scenario with high profit heterogeneity, they differ by a factor of 5.

The number of arriving orders in each planning period is drawn from a negative binomial distribution $NB(\mu, CV)$ with mean μ and coefficient of variation CV. This distribution is commonly used in the literature (Ehrenberg, 1959; Agrawal and Smith, 1996; Quante et al., 2009; Guhlich et al., 2015b). The mean number of arriving orders is varied, resulting in different levels of scarcity of the capacity. We use a CV of 0.5 and 0.75 to represent different levels of demand uncertainty.

For each factor combination from Table 3.2, 30 independent demand scenarios are generated. This leads to 540 instances for a given production system.

		M	Capacity	Order in which machines are visited
5Stage		5	$a_{mt} = 50 \, \forall m$	1, 2, 3, 4, 5
10Stage		10	$a_{mt} = 25 \, \forall m$	1, 2, 3, 4, 5, 6, 7, 8, 9, 10
Bottle		5	$a_{mt} = 50 \, \forall m \neq 3$ $a_{3t} = 40$	1, 2, 3, 4, 5
Reent		4	$a_{mt} = 50 \, \forall m \neq 2$ $a_{2t} = 100$	1, 2, 3, 2, 4
2Prod	Product 1 Product 2	5	$a_{mt} = 75$ for $m = 1, 2$ $a_{mt} = 50$ for $m = 3, 4, 5$	1, 2, 2, 3, 5 1, 1, 2, 4, 5

Table 3.3: Production systems

Note that we use a different order arrival structure than in the SDP (3.4) because in practice, the number of arriving orders within a given planning period is easier to forecast, while the arrival structure in the SDP is a standard formulation from the literature.

To avoid the effects of an empty system, the production system is initially filled with a set of orders at different stages of production. This set of orders is generated using the set of orders present in a production system (work in process) after some planning periods, where the presented revenue management approach was applied with a scarcity of capacity of 100% and a coefficient of variation of 0.5.

Production environments

To investigate the impact of different characteristics of a production system on the performance of the tested algorithms, we consider multiple types of production systems. Specifically, we consider five basic systems to examine the single characteristics of production systems and one practice-inspired production system that has a more complex structure. For the basic production systems, Table 3.3 shows the number of machine groups, the available production capacities, and the sequence in which the machine groups are visited, i.e., the routes of the orders. Each machine group is always required for an entire planning period. At the same time, a machine group might be used in multiple consecutive planning periods.

We consider a

- balanced 5-stage production system *(5Stage)*,
- a balanced 10-stage production system *(10Stage)*,
- a 5-stage production system with a bottleneck *(Bottle)*,
- a 4-stage production system with re-entrant flows *(Reent)*,
- a 5-stage production system with two different product types *(2Prod)*,
- and a more complex production system *(Complex)*.

The complex production system is motivated by a real-world semiconductor backend facility (Ehm et al., 2011). It consists of $M = 23$ machine groups. Two different product types with different routes through the production system are considered. The required capacities and the number of parallel machines for each machine group are given in the appendix in Table A.1. We use the same machines and processing times as described in Ehm et al. (2011). For simplicity, however, we exclude some additional characteristics of the production system, namely splitting orders, batching, sequence-dependent setup times and random machine down times.

In the production systems with two different product types, three different order types exist for each product type. The order arrival rates are the same for each order type in each considered production system.

Algorithms

All algorithms are implemented in C++ using the Gurobi 6.02 solver on a 3.20 GHz Intel Core i7 machine with 32GB of RAM.

The presented revenue management approach *(MS-RM)* is implemented as described in Section 3.5. Bid prices are recomputed at the beginning of planning periods 10, 20, and 30. The convergence of the bid prices is defined as follows: We look at the set of bid prices generated by including the 10 most recent demand scenarios. We add the new bid prices one by one and check whether one of the mean bid prices changes by more than 5 units. If this is not the case, we say that the bid prices have converged. To avoid premature convergence, the minimum number of used demand scenarios is set to 10. The maximum number of considered demand scenarios is 50.

In the numerical experiments, we compare MS-RM with the following benchmark algorithms:

- *First-come-first-serve (FCFS):* The FCFS heuristic is wide spread and used in production planning and control (PPC) because of its simplicity. Here, to avoid excessive backlogging, orders are accepted as long as capacity is available to produce the arriving order without backlogging independent of the profit margin. Orders are released to the shop floor as soon as capacity is available.
- Revenue management approaches from the literature considering either a simplified production system or not using *online* order release decisions
  - *SingleStage:* The SingleStage approach is based on the ATO-RM approach presented in Guhlich et al. (2015b). This algorithm models only one machine group. To apply this algorithm, the production systems considered must thus be aggregated to a single-stage production system. To this end, the capacity, which is determined by the number of parallel machines, of this new machine group is chosen such that the resulting maximum throughput is equal to the maximum throughput of the real system. As in the ATO-RM approach, an order requires the capacity of the only available machine group in each planning period during processing, the number of parallel machines is equal to the throughput multiplied by the (number of planning periods of) lead time. If multiple product types are available, the throughput is multiplied with the mean processing time of the product types weighted by their arrival rate. The available capacity in each planning period is thus 200 for *Bottle*, 248.81 for *Complex*, and 250 for *5Stage*, *10Stage*, *Reent*, and *2Prod*.
  - *SchedArr:* SchedArr uses the same setup as MS-RM, except that the order release date for this order is decided upon order arrival as described by Spengler et al. (2008). An order d is released in the planning period t so that the difference between the marginal profit $prof_{dt}$ and the opportunity costs $opp^{res}(d,t)$ is maximized when this maximum is non-negative. The order release date is fixed from this point on.

Because it is computationally intractable to compute an optimal policy for problems of realistic size, we use an ex-post optimal solution *(PostOpt)* that can be derived from solving the Integer Program consisting of constraints (3.5), (3.7), and (3.9) with objective function (3.8) under the assumption of known future demand. Notably, the results generated with this approach are only an upper bound of the maximum attainable profit because this algorithm has complete demand information, while the other approaches learn the incoming demand only when the orders arrive.

Except for the *Complex* production system, all algorithms operate with a planning period length such that no time discretization defects occur. As known from the literature (Hackman and Leachman, 1989; Stadtler, 2008), an exact model of a production system must consider all points in time at which a processing step may finish. However, using short planning period lengths drastically increases the size of the considered model. In the *Complex* production system, the applied algorithms use a planning period length of 24 hours, although a planning period length of 1 hour would be required to avoid time discretization defects.

Note also that no optimal ex-post solution can be obtained for the *Complex* production system because of the size of the resulting IP. Nevertheless, we derive results for the continuous relaxation of the IP given by $0 \leq x_{dt} \leq 1$, which provides an upper bound for the maximum attainable profit.

In the numerical tests, the production process is deterministic. However, because some of the applied heuristics in this study suffer from modeling defects because they do not model the available capacity exactly, they might compute infeasible schedules. That is, if capacities are overestimated, according to the schedule, orders must wait within the production system, which is forbidden under Assumption 3.3.1. Therefore, the order releases generated by these algorithms are converted to a feasible schedule as follows: Orders are released in the planned planning period if sufficient capacity is available to produce them without waiting; otherwise, released orders are processed in an FCFS manner. This can lead to orders finishing after the planning horizon. For a fair comparison, these orders should not generate positive profit because they use capacity from periods after the planning horizon. Therefore, in this numerical study, orders that finish after the planning horizon earn no profit. However, any costs incurred by these orders are still taken into account.

3.6.2 Numerical results

In this section, we present the results of the numerical study. In particular, we show that the proposed revenue management approach (MS-RM) works well compared to an ex-post optimal solution. Next, we show that aggregating the production system to a single-stage problem does not lead to satisfactory results in production systems with a complex structure. Additionally, the benefit of using online scheduling is described.

The main results can be seen in Figure 3.2, which shows the mean relative difference (in %) of the attained marginal profit over all tested instances compared to PostOpt in all tested production systems for all tested algorithms. Additionally, the 95% confidence intervals are indicated. The following conclusions can be drawn:

- In the tested cases, MS-RM is very close ($\leq 4.4\%$ on average) to an ex-post optimal solution in all considered basic production systems. In the *Complex* production system, the average performance is still good (8.75%), despite time discretization defects. This is a quite impressive result, considering that PostOpt has full demand information, while MS-RM makes online decisions.
- A simple FCFS approach does not lead to satisfactory results, which suggests that exploiting customer differentiation is essential in the tested cases.
- Applying a one-stage-based heuristic leads to satisfactory results in most of the considered basic production systems. However, the results are notably worse in production systems with a more complex structure, e.g., in those production systems with multiple product types.
- Deciding about order releases online significantly increases the marginal profit compared to fixing order release decisions for an order on its arrival, as seen by MS-RM outperforming SchedArr.

Note that the maximum computation time for solving the IP for the order release decisions is set to 5 minutes, which was sufficient to find a feasible, but not necessarily optimal, solution. Thus, the results for MS-RM could lead to slightly better results if this time limit is relaxed.

In the following, we discuss the absolute performance of MS-RM, the impact of aggregating the production system, and the impact of making *online* order release decisions in more detail. To this end, we also examine order fill rates (i.e., the ratio of accepted and available orders) for all order types and the resulting holding/backlog costs per order, which are described in Tables 3.4 and 3.5, respectively. Structural results for PostOpt in the *Complex* production system are unavailable because no integer optimal solution can be computed, as discussed above. In Table 3.5, the costs that were planned by the algorithm are shown in addition to the actual holding/backlog costs because time discretization defects occur in the *Complex* production system. This is abbreviated as Complex (planned).

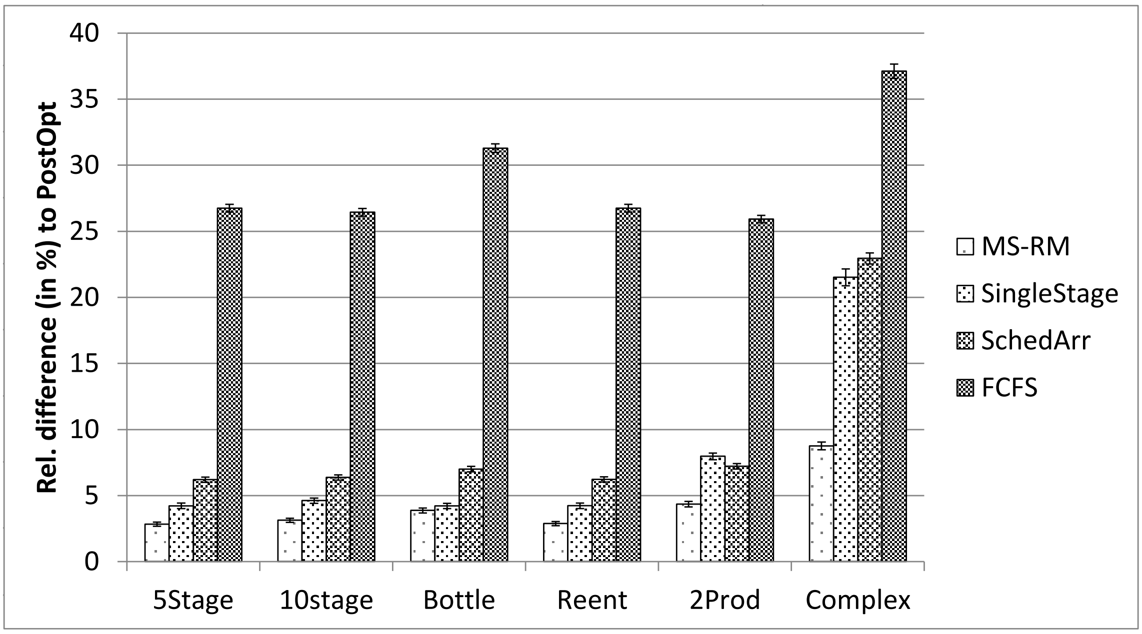


Figure 3.2: Relative differences in attained marginal profit of heuristics compared to an ex-post optimal solution in various production environments

	PostOpt	MS-RM	SingleStage	SchedArr	FCFS
5stage	99.9/99.0/66.0	98.0/95.6/68.7	98.1/95.9/71.3	97.8/95.0/55.3	71.6/71.5/73.2
10stage	99.5/97.5/65.4	97.4/94.1/66.4	97.6/96.0/71.1	97.8/93.1/55.6	71.6/71.5/73.2
Bottle	99.8/93.9/24.0	97.2/93.1/20.1	97.8/94.0/19.9	97.3/89.3/12.6	61.0/60.7/62.4
Reent	99.9/99.0/66.0	97.9/95.6/68.8	98.1/95.9/71.3	97.7/95.0/55.3	71.6/71.5/73.2
2Prod-P1	100.0/99.3/65.5	97.5/94.6/61.1	99.1/97.7/70.8	97.7/94.6/54.4	70.9/71.3/72.8
2Prod-P2	100.0/99.7/64.4	98.1/96.4/63.0	99.0/97.8/71.0	98.0/94.7/54.4	74.4/74.7/74.6
Compl-P1	n.a.	91.1/91.4/74.4	93.5/92.9/77.4	83.2/83.4/55.7	34.7/34.8/35.6
Compl-P2	n.a.	96.6/86.9/39.2	96.4/95.3/62.8	87.1/65.6/29.8	76.4/76.9/76.5

Table 3.4: Overall averages of order fill rates for high/medium/low-value order types

	PostOpt	MS-RM	SingleStage	SingleStagePlanned	SchedArr	FCFS
5stage	2.8/0.1	3.8/1.1	3.9/2.7	4.9/1.1	1.6/5.3	9.8/0.0
10stage	2.0/0.2	2.7/1.4	2.7/3.4	4.2/1.3	1.1/5.7	7.3/0.0
Bottle	3.0/0.4	4.1/2.2	4.3/2.8	5.2/2.0	1.5/5.6	10.1/0.0
Reent	2.8/0.1	3.9/1.2	3.9/2.8	5.0/1.1	1.6/5.4	10.0/0.0
2Prod	2.2/0.2	2.5/2.4	2.0/7.5	4.4/0.6	1.1/6.9	9.8/0.0
Complex	n.a.	1.3/19.0	0.3/31.0	2.2/5.9	0.4/27.9	6.4/0.1
Complex (planned)	n.a.	4.1/8.4	2.2/5.9	2.2/5.9	1.0/21.5	9.0/0.0

Table 3.5: Overall averages of holding/backlog costs per order

		5Stage	10stage	Bottle	Reent	2Prod	Complex
Mean	100%	2.2	2.4	3.5	2.2	3.2	7.1
	110%	2.9	3.2	4.1	3.0	5.0	9.1
	120%	3.3	3.8	4.0	3.5	4.9	10.1
CV	0.5	2.3	2.6	3.0	2.4	3.8	8.8
	0.75	3.3	3.6	4.7	3.3	4.9	8.7
ProfitHet	low	2.8	3.2	4.1	2.9	4.9	8.9
	med	2.9	3.1	3.9	3.0	4.5	8.6
	high	2.8	3.1	3.6	2.8	3.7	8.7

Table 3.6: Overall averages of performance of MS-RM in different production systems

Performance of revenue management approach

In this section, we further investigate the performance of MS-RM compared to an ex-post optimal solution and the FCFS heuristic.

As Figure 3.2 shows, MS-RM clearly outperforms the FCFS heuristic in all production systems, showing that the considered problem cannot be solved using this simple heuristic. The poor performance might be explained by the fact that FCFS does not differentiate between order types as shown in Table 3.4, while PostOpt and MS-RM both accept significantly more high-value orders. Additionally, although FCFS does not create backlog costs by definition, the resulting holding costs are rather high, see Table 3.5.

MS-RM is very close to PostOpt ($\leq 4.4\%$ on average) in all considered production systems except for *Complex*. The good performance is supported by the other performance indicators. The order fill rates for MS-RM and PostOpt are very close although PostOpt still accepts more orders because it has full knowledge of all incoming demand upfront (cf. Table 3.4). In addition, it seems that the order differentiation of MS-RM is not strict enough compared to PostOpt. Also, the holding and backlog costs are on a similar level, as shown in Table 3.5. Again, PostOpt generates lower costs because order release decisions can be performed with the knowledge of all accepted orders.

The performance of MS-RM in the *Complex* production system is slightly worse than for the other production systems because discretization defects occur in modeling this production system, as described in Section 3.6.1. Therefore, capacity is overestimated, leading to 4.5% of the accepted orders finishing after the planning horizon and thus not adding to the earned profit. Additionally, the backlog costs increase from the planned 8.4 per order to an actual 19 per order, as shown in Table 3.5.

Table 3.6 shows the impact of the varied parameters on the performance of MS-RM. Here, the relative difference from an ex-post optimal solution is depicted for varying loads of the system, the coefficient of variation of the incoming demand, and the profit heterogeneity (abbreviated as ProfHet) of the orders.

The performance of MS-RM relative to PostOpt decreases with an increasing load. One reason for this is that there are more orders to decide about and, as a result, more mistakes to make.

MS-RM also performs worse for a higher CV in all production system except for *Complex*. This can be explained by higher fluctuations of demand. Therefore, scenarios in which the computed bid prices are unsuitable for the actual demand arriving occur more frequently. In the *Complex* production system, this factor seems to play a lesser role. One reason for this might be that because the production system is not modeled exactly, in this case, the impact of the time discretization defects is dominant.

It seems that a low profit heterogeneity leads to a slightly better performance. However, the difference is rather small. On the one hand, the impact of having to reject a high-value order is smaller if the profit heterogeneity is low, but on the other hand, it is easier to differentiate between orders in the case of high profit heterogeneity.

Impact of aggregating the production system

In this section, we examine the impact of aggregating the production system to a single-stage system. To this end, we compare the SingleStage approach with MS-RM and the ex-post optimal solution.

Figure 3.2 shows that this approach leads to satisfactory results if the production system is simple. However, if, e.g., multiple product types are available, the modeling defects that stem from the aggregation lead to poor results. Accepted orders that finish after the planning horizon do not contribute to overall profit. For the simple production systems, the percentage of these orders is rather low ($1.0\%, 1.6, \%, 0.5\%$, and 1.0% for *5stage*, *10stage*, *Bottle*, and *Reent*, respectively), while for the production systems with two product types, a significant number of accepted orders cannot be produced within the planning horizon (3.5% for *2Prod* and 13.1% for *Complex*). The same result can be observed when comparing the planned and the actual backlog costs in Table 3.5. While for the simple production systems, the planned and actual realized costs are very similar, they drastically increase in the *2Prod* and *Complex* production systems.

Impact of using online scheduling

In this section, the impact of making order release decisions online relative to fixing order release decisions on order arrival is discussed. To this end, we compare SchedArr with MS-RM and PostOpt.

Figure 3.2 shows that there is a significant benefit to being able to reschedule orders instead of fixing the release dates on arrival. This can be explained by the level of knowledge on the whole set of orders to be scheduled. While PostOpt has full knowledge of all incoming demand and MS-RM at least has the flexibility to change planned order releases if further orders have arrived, SchedArr must determine the order release date for an order immediately on its arrival. In this way, especially in complex production systems, inappropriate order release decisions can lead to resources that can no longer be used. Therefore, SchedArr cannot accept as many orders as MS-RM or PostOpt (cf. Table 3.4).

	5Stage	10Stage	Bottle	Reent	2Prod	Complex
Percentage Preponed	59.5	69.9	58.9	60.2	68.3	48.8
Percentage Delayed	13.6	15.2	11.9	12.5	14.1	18.1

Table 3.7: Percentage of orders rescheduled in MS-RM

Table 3.7 shows that most of the orders are actually rescheduled in MS-RM. Most of the orders are preponed, which might explain why the backlog costs decrease. However, some orders are also delayed, perhaps allowing more orders to be accepted.

As seen in Table 3.5, SchedArr generates high backlog costs, which may be related to the chosen order release policy. Order release dates are chosen such that they minimize the sum of the bid prices for the used resources and the backlog costs. Bid prices tend to be lower for capacity in late planning periods because, for these periods, less capacity is already planned for fixed order releases. This might be a reason why many orders are backlogged using SchedArr.

3.7 Conclusion and further research

In the present paper, we consider the order acceptance and order release decisions in an order-driven multi-stage production environment. To the best of our knowledge, this problem has not yet been studied in the literature. Either a single-stage production system has been considered, or order release decisions have been fixed on order arrival. We describe this problem as an SDP.

To solve instances of realistic size, we develop a bid-price-based revenue management approach where bid prices are computed using an RLP approach. The bid prices are aggregated to estimate the opportunity costs of using resources for specific orders. The order acceptance and order release decisions are made using an LP, which captures the stochastic nature of the demand via the previously computed opportunity costs.

In a numerical study, we show that the presented approach works well for various production systems compared to an ex-post optimal solution. Even for the *Complex* production system, which cannot be modeled exactly due to the high computational burden, the performance is only slightly worse. The application of a simple FCFS heuristic leads to unsatisfactory results in all tested production systems, showing that order differentiation is important in the considered settings. Moreover, the benefit of the presented approach compared to existing approaches from the literature, which either consider aggregated single-stage production or fix order release decisions upon order arrival, is shown.

In the present paper, set-up times are neglected, and processing times are assumed to be deterministic. This is not always the case in real-world production facilities, and it provides a good avenue for further research. Nevertheless, the present paper is a next step in the direction toward enabling the application of revenue management in complex real-world production environments as well.

4 Revenue management in stochastic production systems: A clearing function-based bid price approach

Co-authors:

Moritz Fleischmann
Chair of Logistics and Supply Chain Management, Business School, University of Mannheim, Germany

Lars Mönch
Chair of Enterprise-wide Software Systems, Department of Mathematics and Computer Science, University of Hagen, Germany

Raik Stolletz
Chair of Production Management, Business School, University of Mannheim, Germany

Working paper.

Abstract:

In this paper, we consider the order acceptance and order release decisions of a manufacturer facing order-specific demand. In contrast to previous literature, both demand and production are stochastic. We develop a novel bid-price based revenue management approach to solve this decision problem. Here, the production system is modeled using clearing functions to capture the non-linear inter-dependency of workload and lead times in stochastic production systems. In the existing literature, a common approach to cope with variability in a production system is to introduce fixed, workload-independent waiting times. We show in a numerical study that this fixed lead time approach is clearly outperformed by the new developed approach based on clearing functions presented in this paper.

4.1 Introduction

In this paper, we develop a bid price-based revenue management (RM) approach for the order acceptance and order release decisions of a manufacturer facing stochastic demand in an order-driven, stochastic production environment.

Apart from the classical RM applications in service industries (see Talluri and Van Ryzin, 2004, for an overview), nowadays, RM approaches are also used in manufacturing contexts like make-to-stock (MTS) production systems (e.g. Quante et al., 2009; Meyr, 2009), assemble-to-order (ATO) production systems (e.g. Gao et al., 2012; Guhlich et al., 2015b), and make-to-order (MTO) production systems (see e.g. Hintsches et al., 2010; Chevalier et al., 2015; Guhlich et al., 2015a, and references therein). These papers consider production systems without variability in the production process due to stochastic influences, sequence-dependent set-up times, or batch formation. Therefore, waiting times within the production system are not modeled. In this paper, we consider demand management decisions in production systems that are subject to variability induced by stochastic influences.

As an example for a production environment with variability in the production process, we consider the semiconductor back-end production. Figure 4.1 depicts the supply chain of a typical semiconductor manufacturer consisting of two stages: front-end and back-end production. In the front-end, the integrated circuits are fabricated onto a raw wafer, a thin disc made from silicon, and the finished wafers are stored in a die bank, which serves as the customer order decoupling point. In the back-end production, the functional circuits are cut from the wafers and assembled to finished chips based on customer specific requirements. Then, the finished chips are again tested. This assembly and test process is subject to variability, e.g. due to stochastic influences, sequence-dependent set-up times, and batch building. (see e.g. Mönch et al., 2013).

In this paper, we address the order acceptance and order release decisions of such a manufacturer using a flexible job shop with stochastic effective processing times (e.g. due to stochastic machine breakdowns). Production is order-specific. Demand is stochastic and characterized by industry-wide demand peaks, which cannot be mitigated by expanding available capacities because this is time-consuming and requires high capital investments. In addition, outsourcing is impossible because during a demand peak, all factories work to capacity. Orders differ in their contribution margin, lead time requirements, and capacity requirements. The objective is to maximize the difference of the contribution margin gained

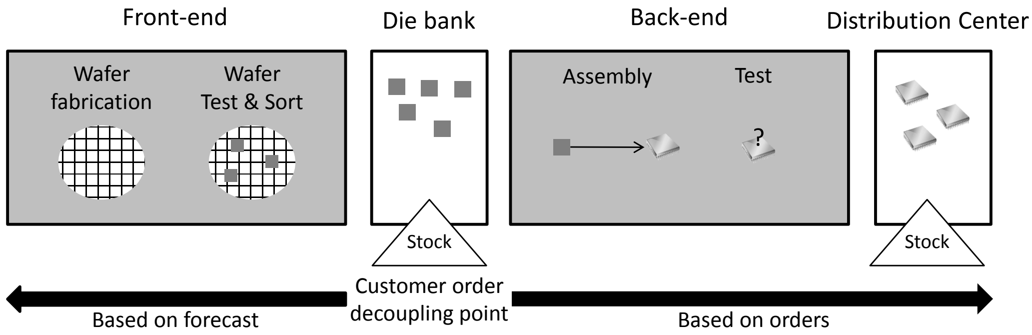


Figure 4.1: Supply chain structure in semiconductor manufacturing (cf. Guhlich et al., 2015b)

by accepted orders and work-in-process *(WIP)* holding costs, finished good inventory *(FGI)* holding costs, and backlog costs.

To tackle this decision problem, we combine bid price-based RM with order release planning based on clearing functions *(CF)*. Bid price approaches are commonly used in RM for order-driven manufacturing. In the existing literature (e.g. Hintsches et al., 2010; Guhlich et al., 2015a), linear programming *(LP)* models with fixed lead times are used to model the production system. This approach is justified in deterministic production systems. However, in systems with stochastic influences, lead times depend non-linearly on the available workload. Therefore, in our approach, we use CFs to model the expected output of a machine given the available workload over a fixed time interval. CF-based LPs have been successfully applied to decide about order releases in production systems with stochastic influences, e.g. in the semiconductor industry (cf. Asmundsson et al., 2006; Kacar et al., 2013).

The main contributions of this paper are as follows:

1. We consider the simultaneous order acceptance and order release decisions in a stochastic production environment with stochastic, order-specific demand. This problem has not yet been studied in literature.
2. We propose a novel bid price approach using CFs to measure the opportunity costs of production capacity.
3. In a numerical study, we show that the presented approach significantly outperforms approaches that cope with variability in the production system by modeling fixed, workload-independent waiting times, which is common practice in order release planning (Missbauer and Uzsoy, 2011).

This paper is organized as follows. In Section 4.2, we review the related literature. Then, we formally define the considered problem in Section 4.3. In Section 4.4, the derivation of CFs is discussed. The bid price-based RM approach is developed in Section 4.5. In Section 4.6, a numerical study evaluates the performance of the presented approach. In Section 4.7, we give a conclusion and show possible avenues for further research.

4.2 Literature review

In this paper, we develop a novel RM approach based on the ideas of bid price-based RM in an order-driven manufacturing context and production planning approaches based on CFs. Therefore, we review literature from these streams. In addition, related literature on admission control and stochastic scheduling approaches is discussed.

Bid price-based RM approaches for order-driven production

Spengler et al. (2007) use a multi-dimensional knapsack formulation to compute bid prices in a deterministic MTO production environment, where all orders are routed via the same path. No scheduling decisions are made because only one start date is available for each order. Hintsches et al. (2010) consider also a deterministic MTO production system. They use a production planning LP assuming fixed lead times to compute bid prices. Volling et al. (2012) use neural networks to update bid prices. Here, capacity is given by a single bottleneck resource with deterministic capacity requirements. Guhlich et al. (2015a) make online order acceptance and order release planning in a deterministic multi-stage no-wait production system using a production planning LP to compute bid prices. Guhlich et al. (2015b) develop a bid price-based RM approach in the context of ATO manufacturing. In addition to production capacity, intermediate materials are considered as scarce resources. They use a production planning LP assuming fixed lead times to model a deterministic single-stage production system.

To sum up, in the existing bid price-based RM approaches for order-driven production, order acceptance and order release decisions are made under stochastic demand. However, all approaches consider deterministic production systems and thus, LPs assuming fixed lead times are used.

Order release planning based on clearing functions

It is known from queuing theory that in stochastic production systems, the relationship between load and output is non-linear (Karmarkar, 1993; Medhi, 2002). This non-linearity is often disregarded and traditional approaches assuming fixed lead times are extended by workload-independent, fixed waiting times (Missbauer and Uzsoy, 2011). In particular, the *flow factor approach* (see e.g. Kacar et al., 2013) determines effective lead time estimates by scaling up the processing time of each processing step by the flow factor, which is defined as the ratio of the cycle time of a product (determined by simulation) and the sum of the raw processing times. However, in recent time, production planning approaches have been developed to cope with load-dependent lead times, see Mula et al. (2006), Pahl et al. (2007), or Missbauer and Uzsoy (2011) for overviews. In this paper, we use CFs and thus, focus on this literature stream in what follows.

Asmundsson et al. (2009) develop a CF model to capture the non-linear relationship between workload and output. To allow for multiple product types, they allocate capacity according to the available WIP of the different product types. Asmundsson et al. (2006) show in a simulation study that this model works well in a semiconductor wafer fabrication facility. The CFs are estimated via visual fit. Kacar et al. (2013) apply this CF approach to a simulation model of a large-scale semiconductor wafer fabrication facility and compare it with benchmark approaches using fixed lead times. Brahimi et al. (2015) integrate order acceptance decisions for known demand in a mixed integer program based on linearized CFs. Missbauer (2009) demonstrates limitations of CF based production planning approaches. Missbauer (2011) criticizes that one-dimensional CFs overestimate the available capacity. Consequently, he proposes a three-dimensional CF based on WIP, the ratio between WIP from the last period and new arriving WIP, and the variability of the arrival process. Aouam and Uzsoy (2015) compare different approaches that consider the production planning of a manufacturer producing to stock under uncertain demand and workload-dependent lead times. Here, no order acceptance decisions are made.

In all of these approaches, the production system may be subject to stochastic influences. However, none of the approaches considers order acceptance decisions under stochastic demand.

Admission control and stochastic scheduling

Apart from RM approaches, admission control approaches (see e.g. Silva et al., 2013; Millhiser and Burnetas, 2013; Kim and Kim, 2014; Mlinar and Chevalier, 2015) consider order acceptance decisions for stochastic demand. However in contrast to the present paper, their release decisions are based on simple dispatching rules. Moreover, the considered production environments are limited to exponentially distributed or deterministic processing times and one or two different product types.

In stochastic scheduling (see e.g. Cai and Zhou, 1997; Wu and Zhou, 2008; Winands et al., 2011; Baker and Altheimer, 2012; Baker, 2014; Framinan and Perez-Gonzalez, 2015), sequencing decisions are made in a stochastic production environment but in their setting, all arriving demand must be satisfied.

Overall, in the existing literature, the order acceptance and order release decisions for stochastic demand in production systems with stochastic influences have not been studied.

4.3 Assumptions

In this section, we define the considered planning problem. In particular, we describe the production system, the production process, the demand, the decisions to be made, and the objective.

The planning horizon is divided in planning periods $t = 1, \ldots, T$, each of length pl. I different order types denoted by $i = 1, \ldots, I$ can be produced.

Assumption 4.3.1 (Production system) *We consider a flexible job shop with M machine groups ($m = 1, \ldots, M$), which consist of identical parallel machines. Infinite buffer capacity is available in front of each machine group. Required raw and intermediate materials are unlimited available.*

Assumption 4.3.2 (Production process) *Orders are processed in a first-come-first-served (FCFS) manner on the machines of each machine group. Each order can only be processed by one machine at a time.*

An order of type i passes L processing steps ($l = 1, \ldots, L$) and for each processing step l, a certain machine group z_{il} is required. The effective processing times may be subject to *stochastic influences. $proc_{il}$ denotes the expected effective processing time of processing step l of order type i. Capacity is measured in the number of parallel machines available in a planning period, and $\xi_{il} := \frac{proc_{il}}{pl}$ is defined as the expected capacity consumption of processing step l of order type i.*

Assumption 4.3.3 (Demand) *Orders arrive according to a known probability distribution. Each order consists of multiple units so that partial acceptance is possible and orders can be fulfilled via partial deliveries over different planning periods. Each order type i has the following characteristics:*

- *(Expected) capacity requirements ξ_{il} at machine group z_{il} for each processing step l*
- *Time span between arrival and due date (δ_i)*
- *FGI holding, WIP holding, backlog cost rate (hc_i, wc_i, bc_i)*
- *Contribution margin $(contr_i)$.*

Assumption 4.3.4 (Decisions) *The company makes the following decisions:*

- Order acceptance: *Upon arrival, the incoming order is immediately either (partially) accepted or rejected.*
- Order release: *At the beginning of each planning period, the fraction to be released of each accepted order is determined. In addition, the sequence in which the orders are released must be determined.*

Orders can be released at the earliest in the first planning period after their arrival.

Assumption 4.3.5 (Objective) *The objective is to maximize the contribution margin gained by accepted orders minus WIP/FGI holding and backlog costs. Orders that finish production after the due date incur backlog costs. Orders finishing before the due date incur FGI holding costs, and WIP holding costs occur for each planning period during which orders are in production.*

4.4 Clearing function-based capacity constraints

In this paper, we combine bid price-based RM with order release planning based on CFs. Therefore, in this section, we describe how CFs are used in production planning LPs in the literature. To this end, we summarize the derivation of the CF-based capacity constraints as described in Asmundsson et al. (2009). The production planning LP that we use in our approach is presented in Section 4.5.1. In addition, we describe how we estimate the parameters for the (linearized) CFs. Table 4.1 summarizes the notation.

4.4.1 Derivation of CF-based capacity constraints

The main idea of using CFs is to model the maximum output of each machine group m as a (non-linear) function f_m of the available workload over a given period of time. To this end, L_{ilt}, the available number of orders of order type i at processing step l in period t is modeled. In addition, X_{ilt} describes the finished production of order type i in processing step l in planning period t, and Z_{ilt} is the fraction of capacity at machine group z_{il} allocated to order type i at processing step l in planning period t.

In the following, we summarize the derivation of the capacity constraints using CFs, which is based on Asmundsson et al. (2009).

The processed workload of order type i at processing step l in period t is limited by the respective CF $f_{z_{il}}$ based on the total available workload at machine group z_{il}, and the total capacity available at this machine group is allocated according to Z_{ilt}. Hence, the capacity constraints are as follows:

$$\xi_{il} \cdot X_{ilt} \leq Z_{ilt} \cdot f_{z_{il}} \left(\sum_{i'=1}^{I} \sum_{l' \in \{k | z_{il} = z_{i'k}\}} \xi_{i'l'} L_{i'l't} \right) \qquad \forall i, l, t \tag{4.1}$$

$$\sum_{i=1}^{I} \sum_{l \in \{l' | z_{il'} = m\}} Z_{ilt} = 1. \qquad \forall m, t \tag{4.2}$$

Constraints (4.2) are required to ensure that the allocation of capacity to the single processing steps is feasible.

The allocation of the capacity of a machine is assumed to be proportional to the mix of WIP at the machine in the planning period:

$$\xi_{il} \cdot L_{ilt} = Z_{ilt} \cdot \sum_{i'=1}^{I} \sum_{l' \in \{k | z_{il} = z_{i'k}\}} \xi_{i'l'} \cdot L_{i'l't}. \qquad \forall i, l, t \tag{4.3}$$

Indices:	
$t, t' = 1, \dots, T$	planning periods
$m = 1, \dots, M$	machine groups
$i = 1, \dots, I$	order types
$l = 1, \dots, L$	processing steps (dependent on order type)
$c \in C(m)$	line segments of the CF for machine group m
Parameters:	
pl	length of a planning period
$proc_{il}$	expected processing time of processing step l of order type i
ξ_{il}	expected capacity consumption of i in processing step l
z_{il}	machine group required for processing step l of order type i
δ_i	time span between arrival and due date for order type i
$contr_i$	contribution margin of order type i
hc_i, wc_i, bc_i	FGI holding/ WIP holding/ backlog costs rate for order type i
d_{it}	demand for i with due date at the beginning of planning period t
d_{it}^{fix}	accepted demand for i with due date at the beginning of pl. period t
f_m	(non-linear) CF for machine group m
$C(m)$	set of line segments of the (linearized) CF at machine group m
α_{cm}	slope of c-th segment of CF for machine group m
β_{cm}	intercept of the c-th segment of CF for machine m
e_{il}	expected number of new planning periods that are entered during processing step l of order type i
W_{il0}, I_{i0}, B_{i0}	initial values for WIP/ FGI/ backlog of order type i
Decision variables:	
$R_{it} = X_{i0t}$	amount of orders of type i released at the beginning of period t
O_{it}	accepted demand of type i with due date at the beginning of period t
State variables:	
X_{ilt}	finished production of order type i in processing step l in planning period t
Z_{ilt}	fraction of capacity at machine z_{il} allocated to order type i at processing step l in period t
L_{ilt}	available number of orders of type i at processing step l in period t
W_{ilt}	WIP of order type i at processing step l at the end of planning period t
I_{it}	FGI of order type i at the beginning of planning period t
B_{it}	backlog from order type i at the beginning of planning period t

Table 4.1: Notation

To be able to use the CF in an LP, we approximate f_m by a piece-wise linear function. That is, the output of each machine group m is limited by the minimum over the set of affine functions $C(m)$. The c-th line segment of the CF for machine group m is defined by the slope α_{cm} and the intercept β_{cm}. Therefore, we define for a measure of WIP W:

$$f_m(W) \approx \min_{c \in C(m)} \{\alpha_{cm} \cdot W + \beta_{cm}\}. \qquad \forall m \tag{4.4}$$

The parameters α and β must be estimated, which is discussed in the next subsection.

Finally combining (4.1), (4.3), and (4.4), we get the capacity constraints

$$\xi_{il} X_{ilt} \leq \alpha_{cz_{il}} \xi_{il} L_{ilt} + \beta_{cz_{il}} Z_{ilt}. \qquad \forall i, l, t, c \in C(z_{il}) \tag{4.5}$$

4.4.2 Estimation of clearing function parameters

Next, we describe how to derive the parameters for the linearized CFs presented above. Apart from a simulation-optimization approach (Kacar and Uzsoy, 2014b), which is very time-consuming, linear regression approaches are mostly used to estimate CFs (Kacar and Uzsoy, 2014a; Kacar et al., 2013).

In this paper, each CF is modeled by five line segments, where the first line segment is fixed as $\alpha_1 = 1$ and $\beta_1 = 0$, and the last line segment has a slope of zero ($\alpha_5 = 0$). The procedure that we use to estimate CFs is as follows:

1. Generate available workload/output data using 100 independent simulation runs with an expected load of about 70% and 85% of the bottleneck machine group.
2. Sort the data points by the available workload. Divide the data points in five sets A_j ($j = 1, \ldots, 5$) of the same size. Apply linear regression to fit each set A_j to an affine function with parameters α_j and β_j for $j = 2, 3, 4$. β_5 is initialized with the expected maximum throughput in a planning period at the corresponding machine group.
3. Use these parameters as input for the iterative Levenberg-Marquardt algorithm (Marquardt, 1963). Apply this algorithm to find vectors α and β with a small squared difference of the data points to the function $f(x) = \min\{x, \alpha_2 \cdot x + \beta_2, \alpha_3 \cdot x + \beta_3, \alpha_4 \cdot x + \beta_4, \beta_5\}$.

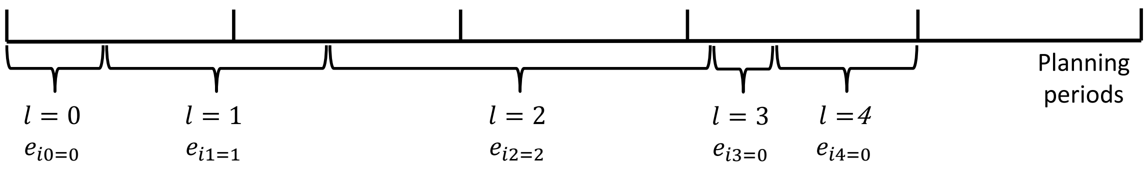


Figure 4.2: Example for different values of e_{il}

4.5 Clearing function-based bid price approach

In this section, we present a heuristic solution approach to the problem described in Section 4.3. We develop a bid price-based RM approach that uses a production planning LP with CF-based capacity constraints to capture the non-linear relationship of workload and lead times in stochastic production systems.

We use this LP to compute bid prices via the randomized linear programming (RLP) approach (Talluri and van Ryzin, 1999). The bid prices are aggregated to estimate the opportunity costs of using capacity. Then, upon arrival of each order, acceptance is decided based on the opportunity costs using the LP. At the beginning of each planning period, the set of released orders is selected, again using the opportunity costs via the LP.

First, we develop the production planning LP using CFs in the capacity constraints as shown in Section 4.4.1. This LP is the basis for the RM approach described afterwards. Next, the computation of opportunity costs is presented. Then, the order acceptance and the order release mechanisms are explained.

4.5.1 CF-based LP

In this subsection, we present the linear constraints that model the production system under the assumption of deterministic demand using linearized CFs as described in Section 4.4.1. Different objective functions are used for bid price computation, order acceptance decisions, and order release decisions, which are defined in Sections 4.5.2, 4.5.3, and 4.5.4, respectively.

In the considered problem, the lead time of an order may exceed the length of a planning period. As we consider individual orders, which are built to order, it is important to know when an order must be released to complete production on time. Therefore, we introduce an offset parameter e_{il}, which approximates the expected number of new planning periods that are entered during processing step l of order type i by assuming that orders do not wait within in the production system. Therefore, e_{il} is defined as the difference between the number of entered planning periods until processing step l is finished and the number of entered planning periods until processing step $l-1$ is finished (under the assumption that orders do not wait within in the production system):

$$e_{il} := \left\lfloor \frac{1}{pl} \sum_{l'=1}^{l} proc_{il'} \right\rfloor - \left\lfloor \frac{1}{pl} \sum_{l'=1}^{l-1} proc_{il'} \right\rfloor . \qquad \forall i, l \qquad (4.6)$$

See Figure 4.2 for an example.

Following this definition, we define the available workload of order type i in processing step l in period t as the sum of W_{ilt-1}, the WIP of order type i at processing step l at the end of planning period $t-1$ and the finished production of order type i at the previous processing step e_{il} planning periods earlier:

$$L_{ilt} := W_{i,l,t-1} + X_{i,l-1,t-e_{il}} . \qquad \forall i, l, t \qquad (4.7)$$

Next, we introduce additional notation to describe the production process. R_{it} is defined as the decision about the number of orders released of order type i at the beginning of planning period t. The FGI and the backlog of order type i at the beginning of planning period t are given by I_{it} and B_{it}, respectively. Let O_{it} denote the decision about the accepted demand of order type i with due date at the beginning of planning period t and let d_{it} and d_{it}^{fix} denote the demand and the already accepted demand for order type i with due date at the beginning of planning period t, respectively.

The following constraints describe the production process:

$$W_{ilt} = W_{i,l,t-1} - X_{ilt} + X_{i,l-1,t-e_{il}} \qquad \forall i, l, t \qquad (4.8)$$

$$I_{it} - B_{it} = I_{i,t-1} - B_{i,t-1} + X_{i,L,t-1-e_{il}} - O_{it} - d_{it}^{fix} \qquad \forall i, t \qquad (4.9)$$

$$O_{it} \leq d_{it} \qquad \forall i, t \qquad (4.10)$$

$$\xi_{il} X_{ilt} \leq \alpha_{cz_{il}} \xi_{il} L_{ilt} + \beta_{cz_{il}} Z_{ilt} \qquad \forall i, l, t, c \in C(z_{il}) \qquad (4.11)$$

$$\sum_{i=1}^{I} \sum_{l \in \{l | z_{il} = m\}} Z_{ilt} = 1 \qquad \forall m, t \qquad (4.12)$$

$$\sum_{t'=0}^{t} R_{it'} \leq \sum_{t'=0}^{t+\delta_i-1} (d_{it'}^{fix} + O_{it'}) \qquad \forall i,t \qquad (4.13)$$

$$B_{iT} = 0 \qquad \forall i \qquad (4.14)$$

$$I_{it}, O_{it}, B_{it}, Z_{ilt}, X_{ilt}, W_{ilt} \geq 0 \qquad \forall i,l,t \qquad (4.15)$$

with $R_{it} = X_{i,0,t}$.

Constraints (4.8) model the WIP of each order type in each processing step in each planning period. The WIP of each order type in each processing step consists of the WIP of the former period subtracted by the finished products in this period and including the new products at this processing step, which correspond either with the finished products of the previous processing step or with new released orders if l is the first processing step. If during this processing step new planning periods have begun according to e_{il}, the intermediate products finished e_{il} planning periods earlier are received. The FGI is modeled by equalities (4.9). Note that I_{it}, B_{it}, O_{it}, and d_{it}^{fix} are related to the beginning of a planning period. Therefore, the inventory of a certain order type is increased by products that have finished their last processing step in the former planning period. The accepted demand for this period and the backlogged orders of the former period are subtracted. If the demand cannot be fully satisfied, the remaining products are added up to the backlog of the next period. Inequalities (4.10) make sure that the accepted demand does not exceed the available demand. Constraints (4.11) and (4.12) describe the available production capacity and are discussed in Section 4.4.1. Constraints (4.13) ensure that orders are only released if a corresponding order is accepted, which is necessary because we assume that production is order-driven. Hence, the cumulative amount of releases in each planning period must not exceed the accepted demand that arrived until the previous planning period (cf. Assumption 4.3.4). t is set off by δ_i periods because the due date of an order is δ_i periods after its arrival. Constraints (4.14) are required to ensure that all accepted demand is satisfied within the planning horizon. Constraints (4.15) describe the non-negativity of the used decision variables.

Compared to the model formulation of Asmundsson et al. (2009), constraints (4.10) and (4.13) are added because we consider order acceptance decisions and the production is order-driven, while in Asmundsson et al. (2009), all demand must be satisfied, and make-to-stock production is assumed. In addition, we introduce the parameter e_{il} to estimate the time between release and finish period of an order. Apart from these differences, compared to the model presented in Brahimi et al. (2015), we allow partial order acceptance, while they model order acceptance using binary variables.

4.5.2 Opportunity costs

To compute bid prices, the problem is modeled by a profit-maximizing LP with constraints (4.8) - (4.15). As is common in literature (Simpson, 1989; Williamson, 1992; Talluri and Van Ryzin, 2004), the primal decisions are discarded and the values of the dual variables corresponding with the capacity restrictions in an optimal solution are used as a measure for the opportunity costs of using these resources. In RLP, demand scenarios $s = 1, \ldots, S$ are generated and the resulting average bid prices are used. Instead of using a fixed number of demand scenarios, we iteratively update the bid prices until the change of the bid prices drops below a certain threshold (c.f. Guhlich et al., 2015b).

The objective function to be maximized is defined as

$$\sum_{i=1}^{I} \sum_{t=1}^{T} \left(contr_i \, O_{it} - \sum_{l=1}^{L} wc_i \, W_{ilt} - hc_i I_{it} - bc_i B_{it} \right). \tag{4.16}$$

This function describes the contribution margin of the accepted orders minus WIP/FGI holding and backlog costs.

The question arises from which constraints the shadow prices should be taken. Similar to Asmundsson et al. (2004), we use the shadow prices of constraints (4.11) (defined as bp_{iltcs} in demand scenario s) as a measure for the marginal costs of one unit of capacity. As at most two line segments of a CF are binding (Asmundsson et al., 2004), we take the maximum over all line segments to derive the opportunity costs for using one unit of capacity for processing step l of order type i in planning period t in demand scenario s:

$$opp_{ilts} := \max_{c \in C(z_{il})} bp_{iltcs}. \qquad \forall i, l, t, s \tag{4.17}$$

With this definition, opportunity costs for different order types using the same machine group may be different. That means, it is possible that a low value order must exceed a lower bid price than a high value order to get accepted. Thus, a low value order might be accepted, while a high value order is rejected although both require the same resources. Therefore, we prefer common opportunity costs for all order types at a given machine group. Numerical tests have shown that taking the minimum over all order types using a machine group leads to appropriate results. Hence, we define by

$$opp_{mts} := \min_{\{i,l|z_{il}=m\}} opp_{ilts} \qquad \forall m,t,s \tag{4.18}$$

the opportunity costs for using machine group m in planning period t in demand scenario s.

Finally, we use the common approach in RLP and take the mean opportunity costs over all demand scenarios:

$$opp_{mt} := \frac{1}{S}\sum_{s=1}^{S} opp_{mts}. \qquad \forall m,t \tag{4.19}$$

4.5.3 Order acceptance decision

In this section, we describe how to make order acceptance decisions based on the opportunity costs computed before. In contrast to former RM approaches (e.g. Guhlich et al., 2015b), we cannot just check feasibility and accept an order if its contribution margin exceeds the opportunity costs of the used resources because it is unclear which capacities are actually used as waiting is allowed within the production system.

Instead, we use an LP that penalizes all production according to the corresponding opportunity costs and allows to accept the order in question. In this way, if the order is accepted, additional production is required, which is penalized according to the opportunity costs. On the other hand, the contribution margin is earned. Therefore, only orders for which the contribution margin exceeds the opportunity costs for the required resources are accepted in an optimal solution of the LP. Another advantage of this method is that it takes into account that the acceptance of an order might generate additional backlog costs for previously accepted orders.

To be precise, if an order of order type i' arrives with due date at the beginning of period t', we set $d_{it} = 1$ if $i = i'$ and $t = t'$, and $d_{it} = 0$, otherwise. Then, we maximize the following objective function

$$contr_{i'} O_{i't'} - \sum_{i=1}^{I}\sum_{t=1}^{T}\left(hc_i I_{it} + bc_i B_{it} + \sum_{l=1}^{L}(wc_i W_{ilt} + opp_{z_{il}t} X_{ilt})\right) \tag{4.20}$$

w.r.t. constraints (4.8) - (4.15). That is, the contribution margin is earned if the new order is accepted but holding and backlog costs are incurred for all orders. Using production capacity is penalized according to the opportunity costs to model the opportunity costs of accepting the order. The order acceptance decision corresponds with the value of O_{it} in an optimal solution of constraints (4.8) - (4.15) maximizing (4.20). Afterwards, $d^{fix}_{i't'}$ is updated accordingly.

4.5.4 Order release decision

At the beginning of each planning period, the set of released orders is selected. To this end, we define a suitable objective function for the constraints (4.8) - (4.15). At this point in time, no order acceptance decisions are to be made, i.e. $O_{it} = d_{it} = 0$ for all i, t.

To take into account future arriving demand, we minimize the holding/backlog costs and the production weighted by opportunity costs (cf. Guhlich et al., 2015b). Note that opportunity costs for capacity in the current planning period can be set to 0 because in the next planning period, this capacity can no longer be used. The objective function to be minimized is

$$\sum_{i=1}^{I}\sum_{t=1}^{T}\left(hc_i I_{it} + bc_i B_{it} + \sum_{l=1}^{L}(wc_i W_{ilt} + opp_{z_{il}t} X_{ilt})\right). \tag{4.21}$$

The decisions for the current planning period are taken, while order release decisions for future periods are discarded.

Finally, the sequence in which orders are released must be determined. Here, the orders to be released are sorted according to nondecreasing due dates. Ties are broken by selecting the order with the higher backlog costs.

4.6 Numerical study

In literature on order release planning as well as in literature on RM, fixed lead time approaches are commonly used to model production systems (Missbauer and Uzsoy, 2011; Guhlich et al., 2015a). In order release planning, it has been shown that CF-based models outperform approaches assuming fixed integer lead times in production systems with variability (see e.g. Kacar et al., 2013). To show that this result can also be reached in an

RM context, in this section, we compare the developed CF-based RM approach *(CF-RM)*[1] presented in Section 4.5 with two RM approaches based on fixed lead times. In particular, the RM approach developed in Guhlich et al. (2015a), which assumes no variability in the production system and thus fixed lead times, serves as basis for these benchmark algorithms. The first benchmark algorithm *(Exp-RM)* completely ignores the variability in the production systems and simply uses the expected processing times as fixed lead times, which is the easiest way to apply existing RM approaches in a production system with stochastic influences. The second benchmark algorithm *(FF-RM)*[2] uses lead times estimated by the flow factor approach. Note that Exp-RM is a special case of FF-RM, where the flow factor is equal to 1, i.e. no waiting time is modeled.

The algorithms are implemented in C++ using a Gurobi 6.0 solver. The simulation environment is provided by AutoSched AP. The numerical tests have been performed on a 3.20 GHz Intel Core i7 machine with 32 GB of RAM.

After presenting the experimental design, we assess the performance in multiple basic production systems to show structural results. Then, we apply the algorithms to a model of a real world like semiconductor back-end facility.

4.6.1 Experimental design

In this section, we describe the simulation environment, the demand settings, and the applied algorithms.

Simulation environment

Order arrivals are simulated and order release plans are computed for all T planning periods. Afterwards, the computed order release plans are executed in a simulation model of the underlying production system. As the considered production systems have stochastic

[1] Bid prices are derived by generating new demand scenarios until the resulting bid prices converge (see Section 4.5.2). That is, until the maximum relative difference of the bid prices changes by less than 5% when including bid prices from new generated demand scenarios. To avoid premature convergence, at least 10 demand scenarios are used. As convergence cannot be guaranteed and bid price computation is computationally demanding, at most 30 demand scenarios are used.

[2] Like in the estimation of CFs (cf. Section 4.4.2), we use simulation data from 100 simulation runs with an expected load of about 70% and 85% of the bottleneck machine to estimate cycle times. To be precise, before each processing step, the orders perform an artificial processing step that requires no capacity and takes the estimated waiting time. Thereby, it is modeled that capacity is required in later planning periods and that the order finishes later.

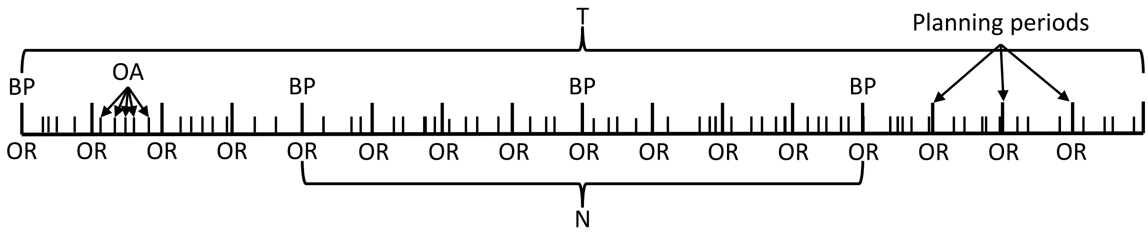


Figure 4.3: Simulation process diagram exemplifying when bid prices are (re-)computed (BP), order acceptance decisions (OA), and order release decisions (OR) are made.

influences, 20 simulation runs with different random seeds are performed for each order release plan.

The general simulation process is depicted in Figure 4.3. At the beginning of each planning period, order release decisions are made for the accepted orders that are not fully released yet. Within each planning period, order acceptance decisions are made for the arriving orders. Due to computational complexity, at the beginning of the planning horizon, we consider only a fixed time window of N planning periods. Every $N/2$ periods, this time window is is extended by $N/2$ planning periods so that again a total of N planning periods is considered. This is done until the maximum amount of overall considered planning periods T is reached. Numerical pre-test have shown that the influence on the results is low if N is chosen sufficiently high. To incorporate information on which orders have been accepted so far, bid prices are recomputed every $N/2$ planning periods taking into account the future N periods.

To mitigate end-of-horizon effects, in the performance evaluation, only orders that finish until $N/2$ planning periods before the end of the planning horizon earn profits. Holding and backlog costs of accepted orders that arrive within this time frame are still taken into account.

Similar to Kacar et al. (2013), the production system is filled with initial WIP computed beforehand to avoid warm-up effects. Five different sets of orders are generated using independent simulation runs of the production system. One of these sets is randomly selected. To allow comparability, the same instances are selected for all algorithms.

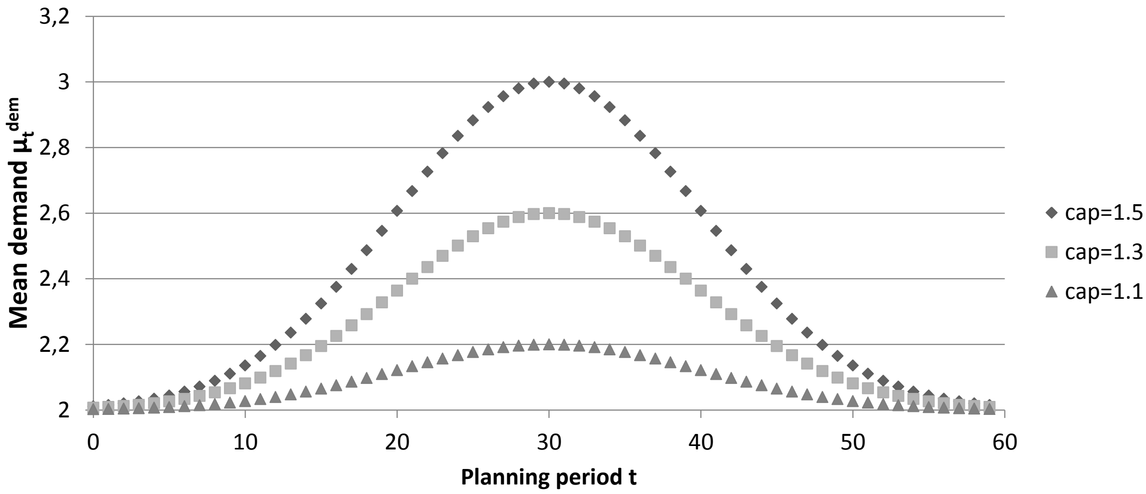


Figure 4.4: Mean demand over time for the three capacity factors with $\mu^{dem} = 2$ and $T = 60$

Demand settings

To model heterogeneous demand, we assume that multiple order types are available. The number of orders of each order type arriving in each planning period t is drawn from a negative binomial distribution with mean μ_t^{dem} and coefficient of variation CV^{dem}. This distribution is commonly used in the literature (Ehrenberg, 1959; Agrawal and Smith, 1996; Quante et al., 2009; Guhlich et al., 2015b). We assume that there is a demand peak in the planning horizon to model that demand exceeds capacity in the short term, which is a typical situation in which RM approaches are applied. Therefore, the demand is time-dependent, and we define

$$\mu_t^{dem} := cap \cdot \mu^{dem} \cdot \exp\left(-\frac{\left(t - \frac{T}{2}\right)^2}{2 \cdot \left(\frac{T}{6}\right)^2}\right) \qquad \forall t, \qquad (4.22)$$

where μ^{dem} depends on the production system and cap determines the height of the demand peak and thus the scarcity of capacity. We use three different capacity factors (1.1, 1.3, and 1.5) and two different values for CV^{dem} (0.75 and 1.0). Figure 4.4 shows an exemplary graph depicting the mean demand over a planning horizon.

To account for the higher willingness to pay of customers with urgent orders, high, medium, and low value order types request a lead time of 2, 3, and 4 periods, respectively. We use two different sets of contribution margins to model different profit heterogeneity settings. Table 4.2 shows the contribution margins for the different order types in these settings.

Factor	Level	Count
Capacity factor cap	$\{1.1; 1.3; 1.5\}$	3
Coefficient of variation CV^{dem} of arriving demand	$\{0.75; 1.0\}$	2
Contribution margin $contr_d$ dependent on order type	$\{(300/200/100);$ $(500/300/100)\}$	2
FGI/WIP holding/Backlog cost rate $hc/wc/bc$ rel. to $contr_d$	$\{(0.5\%/0.25\%/5\%);$ $(1\%/0.5\%/10\%);$ $(2\%/1\%/20\%)\}$	3

Table 4.2: Demand parameters used in the numerical study

Backlog and holding costs are proportional to the contribution margin of an order. We examine three different cost settings, which are also shown in Table 4.2.

For each demand setting, we generate 20 different demand instances according to the given probability distribution. As discussed before, we simulate each computed order release plan with 20 different realizations of processing times. Overall, $3 \cdot 2 \cdot 2 \cdot 3 \cdot 20 \cdot 20 = 14400$ different demand and processing time instances are simulated for each production environment.

4.6.2 Basic production systems

In this subsection, we consider basic production systems with stochastic influences. That is, the effective processing times are stochastic, for example due to manual work or random machine down times. Here, we evaluate, when it is necessary to model a production system using CFs and when applying an approach that assumes fixed lead times is sufficient. To be able to isolate the effects of stochastic variability, we test the algorithms in fairly basic production environments. We first present the tested production environments and show numerical results afterwards.

Production systems

In the considered production systems, three order types with the same capacity requirements are available. Each order consists of $16,000$ single units so that partial acceptance and partial deliveries are possible (cf. Assumption 4.3.3). The effective processing times at each machine are lognormal distributed with mean $\mu^{proc} = 0.9$ seconds and coefficient of variation CV^{proc} for each single unit leading to a mean processing time of 4 hours per order at each visited machine. To test the effect of stochastic variability on the performance of the algorithms, we use three different levels of CV^{proc} (0.5, 1.0, and 1.5). In addition,

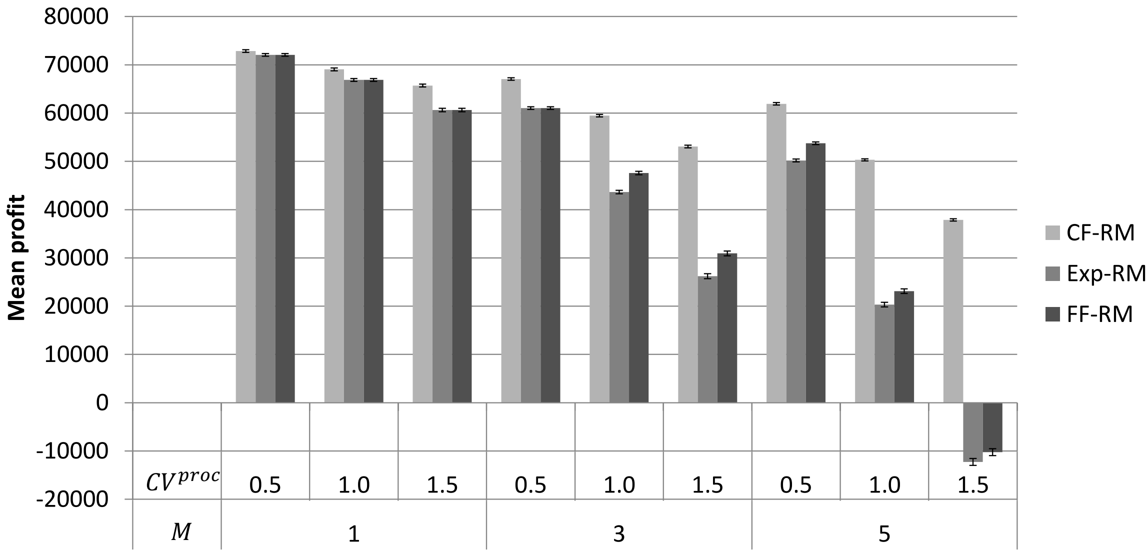


Figure 4.5: Mean profit over all demand scenarios in different production ysystems

we test single-stage, 3-stage, and 5-stage flow shops, i.e. only a single machine is available at each stage. As we consider all combinations of these two factors, overall, nine different production systems are tested. For each production system, CFs and a flow factor are estimated.

The considered planning horizon consists of 60 days divided into $T = 60$ planning periods with a length of $pl = 24$ hours each. We choose $\mu^{dem} = 2$ for each of the three order types, which leads to a fraction of total available workload to total available capacity in the planning horizon of about 104%, 112%, and 120% for the respective factor cap. Bid prices are recomputed every $\frac{N}{2} = 10$ planning periods.

Numerical results

Figure 4.5 shows the mean profits over all tested demand scenarios in the different production systems. In addition, the 95% confidence intervals are indicated. The numerical values are given in Table 4.3. The main results are as follows:

- CF-RM outperforms the benchmark algorithms assuming workload-independent, fixed lead times in all considered production systems and the advantage of CF-RM compared to the benchmark algorithms increases with increasing variability.
- Considering variability in the production process by modeling workload-independent, fixed waiting times only slightly improves the results compared to completely ig-

Prod. system		Mean profit		
M	CV^{proc}	CF-RM	Exp-RM	FF-RM
1	0.5	72849±278	72052±277	72052±277
	1.0	69052±282	66863±291	66863±291
	1.5	65694±296	60629±353	60629±353
3	0.5	67060±262	61029±276	61029±276
	1.0	59465±253	43654±363	47589±344
	1.5	53070±264	26213±533	30934±498
5	0.5	61923±245	50176±304	53754±294
	1.0	50322±223	20329±478	23110±470
	1.5	37877±250	-12266±717	-10241±722

Table 4.3: Mean profits in all tested production systems including 95% confidence intervals

noring stochasticity. It is also required to adapt capacities to make reasonable order acceptance decisions as is done by CF-RM via the CFs.

It can also be seen that overall the obtained profits decrease with increasing variability in the production system. That is, with an increasing number of machines with stochastic influences and with an increasing coefficient of variation. This is intuitive because stochastic variability leads to longer cycle times. Thus, less orders can be accepted. As CF-RM is based on CFs, which consider the non-linear relationship between load and output in stochastic production systems, CF-RM is able to manage increasing variability better than the approaches that assume fixed lead times.

For all single-stage production systems and for the 3-stage production system with low CV^{proc}, the results for Exp-RM and FF-RM coincide because here, the expected lead time including waiting times does not exceed one planning period.

Table 4.4 shows the fill rates, that is, the percentage of accepted orders, for all algorithms in all tested production systems. In general, the algorithms show a similar behavior. They accept most of the high and medium value orders, while low value orders are rejected dependent on the load of the system. With a higher variability, the fill rates for CF-RM decrease, while they remain relatively unchanged for the fixed lead time approaches. Especially Exp-RM accepts almost the same number of orders independent of the variability in the production system. This can be explained by the fact that the different CFs indicate that less orders can be produced in production systems with high variability. Exp-RM does not differentiate at all because the expected values remain the same and for FF-RM, only longer lead times are indicated. However, the number of accepted orders is not so much

Prod. system		FillRates		
M	CV^{proc}	CF-RM	Exp-RM	FF-RM
1	0.5	99.9/98.7/50.6	99.8/98.5/52.5	99.8/98.5/52.5
	1.0	99.9/98.7/46.8	99.8/98.5/53.3	99.8/98.5/53.3
	1.5	99.9/98.2/32.3	99.8/98.5/53.1	99.8/98.5/53.1
3	0.5	99.9/98.3/33.0	99.8/98.5/52.2	99.8/98.5/52.2
	1.0	99.7/96.9/19.8	99.8/98.5/52.2	99.6/98.4/47.7
	1.5	99.4/93.4/7.3	99.8/98.5/52.2	99.6/98.4/47.7
5	0.5	99.7/96.9/16.3	99.8/98.5/52.1	99.7/98.4/47.3
	1.0	98.8/88.4/6.2	99.8/98.5/52.1	99.1/97.7/50.8
	1.5	98.1/81.8/1.3	99.8/98.5/52.1	97.9/97.2/53.3

Table 4.4: FillRates (High/Medium/Low value orders) in all tested production systems

Prod. system		Costs		
M	CV^{proc}	CF-RM	Exp-RM	FF-RM
1	0.5	1.5/1.3/12.3	1.3/1.4/14.8	1.3/1.4/14.8
	1.0	1.4/1.9/20.3	1.1/2.1/27.4	1.1/2.1/27.4
	1.5	1.7/2.3/26.0	1.1/3.0/44.1	1.1/3.0/44.1
3	0.5	0.6/2.7/22.9	0.3/3.3/42.8	0.3/3.3/42.8
	1.0	0.5/3.8/41.6	0.2/5.7/89.0	0.2/5.4/78.5
	1.5	0.5/4.8/58.0	0.1/8.1/136.5	0.2/7.7/124.5
5	0.5	0.3/3.6/34.1	0.1/4.9/72.6	0.1/4.7/62.6
	1.0	0.2/5.3/63.3	0.0/9.0/153.8	0.0/8.9/146.7
	1.5	0.1/7.4/103.8	0.0/13.4/242.3	0.0/13.4/236.7

Table 4.5: Costs (FGI/WIP/backlog costs per accepted order) in all tested production systems

affected because the required capacities remain unchanged based on the mean effective processing times.

Table 4.5 shows the resulting cost components for the algorithms in all tested production systems. The WIP holding and backlog costs per order increase for all approaches with increasing variability, which is in line with expectations. However, the increase is a lot higher for the approaches assuming fixed lead times than for CF-RM because here, the non-linear dependency of load and lead times is considered and thus, as seen before, the fill rates are reduced in case of higher variability. The high fill rates for Exp-RM and FF-RM lead to rather high backlog costs and imply that these approaches probably accept too

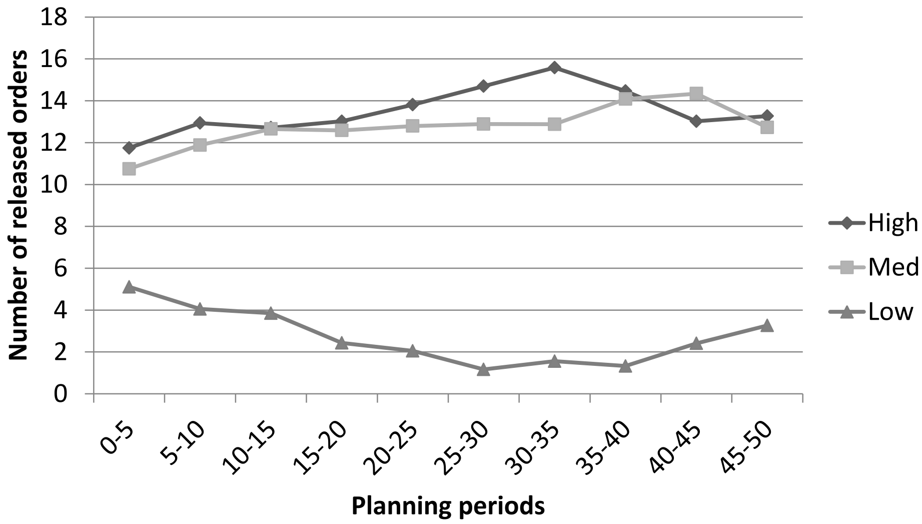


Figure 4.6: Mean number of released orders of CF-RM of different order types over the planning horizon in all production systems

much orders. Therefore, the lead times used in Exp-RM and FF-RM are not valid anymore because they were estimated under the assumption of a lower load.

From these results, we conclude that the applicability of the flow factor approach in this problem setting is limited. In the order release decision literature, the approximate load of a production system is given because the set of orders to be released is fixed. Therefore, the expected lead time of orders can be estimated. However, in the considered problem setting, orders can be rejected and thus, the load of the system and therefore also the cycle times depend on the order acceptance decisions. In order to make reasonable order acceptance decisions, the required capacities must also be adapted, which is modeled by CF-RM via the CFs but ignored in the flow factor approach.

The influence of different demand scenarios on the performance of RM approaches has already been studied in literature, see e.g. Guhlich et al. (2015a). Therefore, we do not discuss this topic in detail.

Finally, we examine the influence of the demand peak on the order release decisions over time. Figure 4.6 shows the mean number of released orders of CF-RM over the planning horizon in all production systems. It can be seen that during the demand peak the number of released low value orders decreases. At the same time, the number of high value orders released increases because more high value orders are accepted during the peak. The release of the medium value orders is delayed slightly after the peak. The graph looks

m	Name	m	Name	m	Name	m	Name
1	Dicing	7	WireBonding	13	Singulation	19	SMT
2	Lamination	8	Cleaning	14	Load / Unload	20	SolderReflow
3	BacksideGrinding	9	Molding	15	BurnIn	21	ModuleTest
4	DieBonding	10	PostMoldCure	16	LoSpeedTest	22	ApplicationTest
5	Curing	11	BallAttach	17	HiSpeedTest	23	Q Gate
6	Cleaning_MD	12	Reflow	18	MSP		

Table 4.6: Machine groups in the semiconductor back-end facility

similar for Exp-RM and FF-RM with the difference that in general notably more low value orders are accepted and thus also released with these algorithms.

4.6.3 Semiconductor back-end facility

To confirm that the results derived in the last section hold also in a more real world like setting, we now consider the model of a back-end semiconductor facility already used in Ehm et al. (2011). After describing the production environment, we present numerical results.

Production system

The production system consists of 23 machine groups, see Table 4.6 for a brief explanation of their purpose. Six order types are available, where order types 1, 2, and 3 and order types 4, 5, and 6 have the same routing as shown in Figure 4.7. Table B.2 in the appendix depicts the required machine group m and the required processing time for a single chip in each processing step l. Each order relates to a lot of 16000 chips. At machine group 7, the order is split in 3 lots of equal size for processing. Machine group 15 is a batching machine, i.e. a batch is built of up to 3000 chips, which are then produced in parallel. The processing time at this machine group relates to the processing time for a whole batch. Machine group 16 is subject to random down times with exponential distributed mean time to failure of 4 hours and exponential distributed mean time to repair of 0.75 hours. The number of identical, parallel machines for each machine group and the sequence dependent setup times are also given in Table B.2 in the appendix. Orders are processed in a FCFS manner on each machine group. In case of required set-ups, orders that require the set-up that is already installed

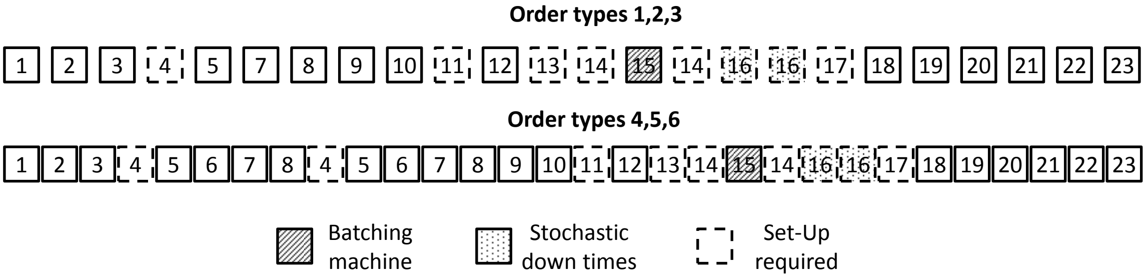


Figure 4.7: Sequence in which machines are visited

	CF-RM	Exp-RM	FF-RM
Mean profit	613511	533974	556506
Fill rates - Order types 1,2,3	100/98.7/28.3	99.9/99.2/59.4	99.8/99.2/55.9
Fill rates - Order types 4,5,6	99.8/95.3/10.2	99.8/98.2/49.1	99.8/99.0/44.0
Costs	0.6/5.1/22.2	0.3/5.6/45.0	0.4/5.4/39.8

Table 4.7: Simulation results for semiconductor back-end facility

on the machine are processed with higher priority. Exact specifications can also be found online at `http://p2schedgen.fernuni-hagen.de/index.php?id=242`.

We consider a planning horizon of 90 days divided into 30 planning periods with a length of 72 hours each. We use longer planning periods in this model because the lead times of the products are also notably longer. We choose $\mu^{dem,1} = 23$ for the order types 1, 2, and 3 and $\mu^{dem,2} = 11.5$ for the order types 4, 5, and 6, which leads to a fraction of total available workload to total available capacity in the planning horizon of about 106%, 112%, and 118% for the respective factor cap. Bid prices are recomputed already every $\frac{N}{2} = 5$ planning periods because in this scenario, the arrival rate is much higher than in the basic production systems so that information on accepted orders must be updated more frequently.

Numerical results

Table 4.7 shows the simulation results for the semiconductor back-end facility. They are similar to the results derived in the basic production systems. CF-RM clearly outperforms the benchmark algorithms that assume fixed lead times. The main difference is in the fill rates, which lead to high backlog costs for the algorithms that assume fixed lead times. Therefore, we conclude that also in a more real-world like production setting, the application of a RM approach based on CFs is beneficial.

4.7 Conclusion and further research

In this paper, we consider the order acceptance and order release decisions of a manufacturer using a production system with stochastic influences and facing stochastic demand. We develop a novel bid price-based RM approach using CFs to capture the non-linear relationship of workload and lead times. This algorithm combines the ideas of bid price-based RM with order release planning based on CFs. Bid prices are computed based on an LP modeling the production system with CFs. The order acceptance and order release decisions are based on a similar LP using the previously computed bid prices in the objective function.

In a numerical study, we test nine production systems with different levels of variability and one model of a real-world like semiconductor back-end facility. We show that simply replacing stochastic processing times by their expected values leads to rather poor results in case of stochastic variability in the production system. In addition, using a common approach in practice and literature, which integrates simulated waiting time estimates, only slightly improves the results.

In the present paper, we assume that during the planning horizon neither updated demand data nor feedback from the execution level is available. If updated demand forecasts are available, they can easily be used in the bid price computation. However, it would be interesting to examine the effects of using more advanced forecast evolution models (see e.g. Albey et al., 2015). In addition, extending the presented approach to use updated production data provides a promising avenue for further research. In the present paper, it is possible that the LP (4.8) - (4.15) is rendered infeasible if the actual processing times exceed the expected values. Also, in the current paper, unlimited availability of intermediate materials is assumed. Considering this factor might be an interesting future research direction. Furthermore, the proposed approach could be tested in larger production systems with more order types that have different routings to see if the observed results still hold.

5 Conclusions and Outlook

5.1 Conclusions

This thesis considers bid price-based revenue management approaches to make the demand management decisions of a manufacturer in different order-driven production environments under the assumption of stochastic, heterogeneous demand and scarce capacities.

It can be concluded that the concept of bid price-based revenue management approaches is applicable in many different production environments. The general methodology is similar in all three articles. The production system is modeled as a linear program and bid prices are derived from the capacity constraints. Then, these bid prices are used to decide about order acceptance and order releases. Differences stem from the considered production systems and assumptions on the demand fulfillment processes.

In the first article, a manufacturer using the assemble-to-order production principle is considered. Here, a deterministic single-stage assembly process is assumed, where both intermediate materials and production capacity are scarce resources. The second and third article consider manufacturers using the make-to-order production principle without scarce intermediate materials. In particular, the second article analyzes a deterministic no-wait multi-stage production system, and the third article examines a stochastic multi-stage production system.

Accordingly, the linear programs describing the production systems are different. In the first and second article, deterministic capacity consumption under the assumption of no waiting in the production system is modeled. For the first article, additional inventory balance equations for the intermediate materials are included in the model. In the third article, the production system is modeled using clearing functions to consider waiting times during the production process, which is important in the case of stochastic processing times. Here, work in process inventory is explicitly modeled for each planning period.

The aggregation of bid prices to opportunity costs for accepting an order differs in the three articles due to the different capacity constraints in the respective linear programs. In addition, in the first article, a due date is quoted, which must be kept. Here, backlog costs depend only on the quoted due date, while holding costs depend on the actual delivery date. In the other two articles, backlog and holding costs depend only on the actual delivery date leading to different decisions to be made.

Although no optimal policies could be derived for instances of realistic size, the presented heuristic approaches perform well. In the first two articles, the developed heuristics are compared with ex-post optimal solutions. The results suggest that the computed solutions are very close to optimal solutions. This is noteworthy because in the ex-post optimal solution, information on all incoming demand is available, while the revenue management approaches only have information on the probability distribution of the incoming demand. In addition, it is shown that the developed approaches significantly increase profits compared to simple decision rules (FCFS) and existing approaches from literature.

5.2 Further research directions

This thesis provides insights how to apply revenue management for order-driven manufacturing. Nevertheless, there are still some possible future research areas.

In the first article, intermediate materials are considered, while in the other two articles they are disregarded. It might be interesting to combine the first and the third article to model a stochastic multi-stage production process with scarce intermediate materials. Here, intermediate materials could be required at multiple stages of the production. It might even be possible to model the production of the intermediate materials.

If intermediate materials are required, the company might have the opportunity to use high value products to satisfy low value demand. In addition, customers might buy substitute products if their order has been rejected. These factors are disregarded in this thesis.

Furthermore, in this thesis it is assumed that demand follows a known probability distribution. However, in practice, it is difficult to determine these probability distributions. Therefore, the robustness of the demand management process with regard to uncertainty in the probability distribution of the demand should be examined. In a next step, one could think of approaches that do not use probability distributions but more rough estimates on incoming demand, which are easier to predict.

All articles assume a finite planing horizon although many products have in principle an infinite selling season. Taking a rolling planning horizon approach, more information becomes available over time and the question is how to incorporate it in the presented approaches. Incorporating updated forecast information can be done by using these forecasts in the bid price update. However, it is interesting to see what effect a more accurate modeling of demand forecasts has on the results. If the production system cannot be modeled exactly, like in a stochastic production system, information from the actual production process might be updated. It is unclear how to incorporate this in the given approaches because the linear programs modeling the production processes might become infeasible if production takes longer than planned.

We found that solving the stochastic dynamic programs describing the problems considered in this thesis is computationally intractable because of the high dimensionality of the resulting state space. Therefore, we developed heuristic approaches to solve instances of realistic size. It would be interesting to find out what makes these problems so hard to solve. Is it for example possible to solve the stochastic dynamic programs if the scheduling decisions are fixed?

Appendix A

Product type 1			Product type 2			Number machines	
i	m	cap_{dim}	i	m	cap_{dim}	m	a_{im}
0	0	0.04	0	0	0.04	0	7
0	1	0.04	0	1	0.04	1	7
0	2	0.04	0	2	0.04	2	7
0	3	0.13	0	3	0.13	3	21
0	4	0.33	0	4	0.33	4	42
0	6	0.29	0	5	0.13	5	14
0	7	0.13	0	6	0.29	6	45
1	7	0.13	1	3	0.13	7	28
1	8	0.04	1	4	0.33	8	7
1	9	0.29	1	5	0.13	9	35
1	10	0.04	1	6	0.17	10	7
1	11	0.29	1	7	0.25	11	35
1	12	0.08	2	6	0.13	12	10
1	13	0.04	2	7	0.25	13	7
1	14	0.08	2	8	0.04	14	56
2	13	0.04	2	9	0.29	15	35
2	14	0.92	2	10	0.04	16	14
2	15	0.04	2	11	0.25	17	7
3	15	0.62	3	11	0.04	18	3
3	16	0.13	3	12	0.08	19	3
3	17	0.04	3	13	0.04	20	7
3	18	0.04	3	14	0.83	21	87
3	19	0.04	4	14	1	22	3
3	20	0.04	5	13	0.04		
3	21	0.08	5	15	0.67		
4	21	0.67	5	16	0.13		
4	22	0.04	5	17	0.04		
			5	18	0.04		
			5	19	0.04		
			5	20	0.04		
			6	21	0.75		
			6	22	0.04		

Table A.1: Capacity requirements for both product types ($cap_{dim} = 0$ for all i, m not shown in the table) and number of machines per machine group in *Complex* production system

Appendix B

Order types 1,2,3				Order types 4,5,6				Num mach.		Setups		
l	m	time[s]	setup config.	l	m	time[s]	setup config.	m	num	current config.	next config.	time[min]
1	1	0.316	-	1	1	0.316	-	1	7	ind	S1	90
2	2	0.316	-	2	2	0.316	-	2	7	ind	S2	15
3	3	0.316	-	3	3	0.316	-	3	7	ind	S3	120
4	4	0.6	S1	4	4	0.6	S9	4	21	ind	S4	30
5	5	1.898	-	5	5	1.898	-	5	42	ind	S5	30
6	7	4.74	-	6	6	0.632	-	6	14	S7	S6	90
7	8	1.266	-	7	7	4.74	-	7	45	S14	S6	180
8	9	0.316	-	8	8	1.266	-	8	28	S15	S6	270
9	10	1.562	-	9	4	0.6	S9	9	7	S6	S7	60
10	11	0.316	S2	10	5	1.898	-	10	35	S14	S7	240
11	12	1.582	-	11	6	0.632	-	11	7	S15	S7	180
12	13	0.474	S3	12	7	4.74	-	12	35	ind	S8	90
13	14	0.316	S4	13	8	1.266	-	13	10	ind	S9	90
14	15	86400	-	14	9	0.316	-	14	7	ind	S10	15
15	14	0.316	S5	15	10	1.562	-	15	56	ind	S11	120
16	16	1.898	S6	16	11	0.316	S10	16	35	ind	S12	30
17	16	1.898	S7	17	12	1.582	-	17	14	ind	S13	30
18	17	0.632	S8	18	13	0.474	S11	18	7	S6	S14	180
19	18	0.316	-	19	14	0.316	S12	19	3	S7	S14	270
20	19	0.158	-	20	15	158400	-	20	3	S15	S14	90
21	20	0.158	-	21	14	0.316	S13	21	7	S6	S15	240
22	21	0.316	-	22	16	1.898	S14	22	87	S7	S15	180
23	22	3.956	-	23	16	1.898	S15	23	3	S14	S15	60
24	23	0.158	-	24	17	0.632	S16			ind	S16	90
				25	18	0.316	-					
				26	19	0.158	-					
				27	20	0.158	-					
				28	21	0.316	-					
				29	22	3.956	-					
				30	23	0.158	-					

Table B.2: Processing steps with required setup configurations for all order types, number of parallel machines, and setup times in the considered semiconductor back-end facility. If the setup time is independent of the former installed setup, this is indicated by “ind”.

Bibliography

Adelman, D. (2007). Dynamic Bid Prices in Revenue Management. *Operations Research*, 55(4):647–661.

Agrawal, N. and Smith, S. A. (1996). Estimating negative binomial demand for retail inventory management with unobservable lost sales. *Naval Research Logistics*, 43(6):839–861.

Albey, E., Norouzi, A., Kempf, K. G., and Uzsoy, R. (2015). Demand Modeling With Forecast Evolution: An Application to Production Planning. *IEEE Transactions on Semiconductor Manufacturing*, 28(3):374–384.

Aouam, T. and Uzsoy, R. (2015). Zero-order production planning models with stochastic demand and workload-dependent lead times. *International Journal of Production Research*, 53(6):1661–1679.

Asmundsson, J., Rardin, R. L., Turkseven, C. H., and Uzsoy, R. (2009). Production planning with resources subject to congestion. *Naval Research Logistics*, 56(2):142–157.

Asmundsson, J., Rardin, R. L., and Uzsoy, R. (2004). Tractable Nonlinear Capacity Models for Production Planning Part I : Modelling and Formulations. Technical report, Laboratory for Extended Enterprises at Purdue School of Industrial Engineering, West Lafayette, IN 47907-1287.

Asmundsson, J., Rardin, R. L., and Uzsoy, R. (2006). Tractable nonlinear production planning models for semiconductor wafer fabrication facilities. *IEEE Transactions on Semiconductor Manufacturing*, 19(1):95–111.

Baker, K. R. (2014). Minimizing earliness and tardiness costs in stochastic scheduling. *European Journal of Operational Research*, 236(2):445–452.

Baker, K. R. and Altheimer, D. (2012). Heuristic solution methods for the stochastic flow shop problem. *European Journal of Operational Research*, 216(1):172–177.

Barut, M. and Sridharan, V. (2005). Revenue Management in Order-Driven Production Systems. *Decision Sciences*, 36(2):287 – 317.

Benjaafar, S. and ElHafsi, M. (2006). Production and Inventory Control of a Single Product Assemble-to-Order System with Multiple Customer Classes. *Management Science*, 52(12):1896–1912.

Brahimi, N., Aouam, T., and Aghezzaf, E.-H. (2015). Integrating order acceptance decisions with flexible due dates in a production planning model with load-dependent lead times. *International Journal of Production Research*, 53(12):3810–3822.

Brown, A., Lee, H., and Petrakian, R. (2000). Xilinx improves its semiconductor supply chain using product and process postponement. *Interfaces*, 10(5):65–80.

Cai, X. and Zhou, S. (1997). Scheduling stochastic jobs with asymmetric earliness and tardiness penalties. *Naval Research Logistics*, 44(6):531–557.

Charnsirisakskul, K., Griffin, P. M., and Keskinocak, P. (2006). Pricing and scheduling decisions with leadtime flexibility. *European Journal of Operational Research*, 171(1):153–169.

Chen, C.-y., Zhao, Z., and Ball, M. O. (2001). Quantity and due date quoting available to promise. *Information Systems Frontiers*, 3(4):477–488.

Chen, C.-y., Zhao, Z., and Ball, M. O. (2002). A model for batch advanced available-to-promise. *Production and Operations Management*, 11(4):424–440.

Cheng, T., Gao, C., and Shen, H. (2011). Production planning and inventory allocation of a single-product assemble-to-order system with failure-prone machines. *International Journal of Production Economics*, 131(2):604–617.

Chevalier, P., Lamas, A., Lu, L., and Mlinar, T. (2015). Revenue management for operations with urgent orders. *European Journal of Operational Research*, 240(2):476–487.

Easton, F. and Moodie, D. R. (1999). Pricing and lead time decisions for make-to-order firms with contingent orders. *European Journal of Operational Research*, 116(2):305–318.

Ehm, H., Wenke, H., Mönch, L., Ponsignon, T., and Forstner, L. (2011). Towards a supply chain simulation reference model for the semiconductor industry. *Proceedings of the 2011 Winter Simulation Conference*, pages 2119–2130.

Ehrenberg, A. S. C. (1959). The pattern of consumer purchases. *Applied Statistics*, 8(1):26–41.

ElHafsi, M. (2009). Optimal integrated production and inventory control of an assemble-to-order system with multiple non-unitary demand classes. *European Journal of Operational Research*, 194(1):127–142.

Framinan, J. M. and Perez-Gonzalez, P. (2015). On heuristic solutions for the stochastic flowshop scheduling problem. *European Journal of Operational Research*, 246(2):413–420.

Gallego, G., Iyengar, G., Phillips, R., and Dubey, A. (2004). Managing flexible products on a network. Technical Report TR-2004-01, Columbia University.

Gallego, G. and Phillips, R. (2004). Revenue Management of Flexible Products. *Manufacturing & Service Operations Management*, 6(4):321–337.

Gallien, J., Tallec, Y. L., and Schoenmeyr, T. (2004). A model for make-to-order revenue management. Working paper, MIT, Cambridge.

Gao, L., Xu, S. H., and Ball, M. O. (2012). Managing an available-to-promise assembly system with dynamic short-term pseudo-order forecast. *Management Science*, 58(4):770–790.

Gönsch, J., Koch, S., and Steinhardt, C. (2014). Revenue management with flexible products: The value of flexibility and its incorporation into DLP-based approaches. *International Journal of Production Economics*, 153:280–294.

Guhlich, H., Fleischmann, M., Mönch, L., and Stolletz, R. (2015a). Revenue management in a multi-stage make-to-order production system. Working paper, University of Mannheim.

Guhlich, H., Fleischmann, M., and Stolletz, R. (2015b). Revenue management approach to due date quoting and scheduling in an assemble-to-order production system. *OR Spectrum*, 37(4):951–982.

Hackman, S. and Leachman, R. (1989). A general framework for modeling production. *Management Science*, 35(4):478–495.

Hall, N. G. and Sriskandarajah, C. (1996). A Survey of Machine Scheduling Problems with Blocking and No-Wait in Process. *Operations Research*, 44(3):510–525.

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

Hintsches, A., Spengler, T. S., Volling, T., Wittek, K., and Priegnitz, G. (2010). Revenue Management in Make-To-Order Manufacturing: Case Study of Capacity Control at ThyssenKrupp VDM. *Business Research*, 3(2):173–190.

Kacar, N. B., Mönch, L., and Uzsoy, R. (2013). Planning Wafer Starts Using Nonlinear Clearing Functions: A Large-Scale Experiment. *IEEE Transactions on Semiconductor Manufacturing*, 26(4):602–612.

Kacar, N. B. and Uzsoy, R. (2014a). A comparison of multiple linear regression approaches for fitting clearing functions to empirical data. *International Journal of Production Research*, 52(11):3164–3184.

Kacar, N. B. and Uzsoy, R. (2014b). Estimating Clearing Functions for Production Resources Using Simulation Optimization. *IEEE Transactions on Automation Science and Engineering*, 12(2):539–552.

Karmarkar, U. S. (1993). Manufacturing lead times, order release and capacity loading. In *Handbooks in Operations Research and Management Science*, volume 4, chapter 6, pages 287–329.

Kim, B. and Kim, J. (2014). Optimal admission control for two station tandem queues with loss. *Operations Research Letters*, 42(4):257–262.

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

Kniker, T. and Burman, M. (2001). Applications of revenue management to manufacturing. In *Third International Conference on Design and Analysis of Manufacturing Systems*, pages 299–308.

Kolisch, R. (2001). *Make-to-order assembly management*. Springer.

Kuhn, H. and Defregger, F. (2005). Revenue management for a make-to-order company with limited inventory capacity. *OR Spectrum*, 29(1):137–156.

Li, X., Wang, J., and Sawhney, R. (2012). Reinforcement learning for joint pricing, lead-time and scheduling decisions in make-to-order systems. *European Journal of Operational Research*, 221(1):99–109.

Lin, J. T., Hong, I.-H., Wu, C.-H., and Wang, K.-S. (2010). A model for batch available-to-promise in order fulfillment processes for TFT-LCD production chains. *Computers & Industrial Engineering*, 59(4):720–729.

Liu, T. (2005). *Revenue Management Models in the Manufacturing Industry*. PhD thesis, MIT.

Marquardt, D. W. (1963). An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2):431–441.

Medhi, J. (2002). *Stochastic models in queueing theory*. Academic Press.

Meyr, H. (2009). Customer segmentation, allocation planning and order promising in make-to-stock production. *OR Spectrum*, 31(1):229–256.

Millhiser, W. P. and Burnetas, A. N. (2013). Optimal admission control in series production systems with blocking. *IIE Transactions*, 45(10):1035–1047.

Missbauer, H. (2009). Models of the transient behaviour of production units to optimize the aggregate material flow. *International Journal of Production Economics*, 118(2):387–397.

Missbauer, H. (2011). Order release planning with clearing functions: A queueing-theoretical analysis of the clearing function concept. *International Journal of Production Economics*, 131(1):399–406.

Missbauer, H. and Uzsoy, R. (2011). Optimization Models of Production Planning Problems. In *Planning Production and Inventories in the Extended Enterprise*, volume 151, chapter 16, pages 393–436.

Mlinar, T. and Chevalier, P. (2015). Dynamic admission control for two customer classes with stochastic demands and strict due dates. *International Journal of Production Research*, (2014):1–18.

Modarres, M. and Sharifyazdi, M. (2009). Revenue management approach to stochastic capacity allocation problem. *European Journal of Operational Research*, 192(2):442–459.

Mönch, L., Fowler, J. W., and Mason, S. J. (2013). *Production Planning and Control for Semiconductor Wafer Fabrication Facilities*, volume 52. Springer New York.

Mula, J., Poler, R., García-Sabater, J., and Lario, F. (2006). Models for production planning under uncertainty: A review. *International Journal of Production Economics*, 103(1):271–285.

Pahl, J., Voß, S., and Woodruff, D. L. (2007). Production planning with load dependent lead times: An update of research. *Annals of Operations Research*, 153(1):297–345.

Petrick, A., Gönsch, J., Steinhardt, C., and 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., and 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*, 93-94:239–252.

Quante, R., Fleischmann, M., and Meyr, H. (2009). *A Stochastic Dynamic Programming Approach to Revenue Management in a Make-to-Stock Production System*. Erasmus Research Institute of Management, Erasmus University.

Silva, D. F., Zhang, B., and Ayhan, H. (2013). Optimal admission control for tandem loss systems with two stations. *Operations Research Letters*, 41(4):351–356.

Simpson, R. W. (1989). Using network flow techniques to find shadow prices for market and seat inventory control. Technical report.

Song, J. and Zipkin, P. (2003). Supply chain operations: Assemble-to-order systems. *Handbooks in Operations Research and Management Science*, 11:561–596.

Spengler, T., Rehkopf, S., and Volling, T. (2007). Revenue management in make-to-order manufacturing - an application to the iron and steel industry. *OR Spectrum*, 29(1):157–171.

Spengler, T. S., Volling, T., and Hintsches, A. (2008). Integration von Revenue Management Konzepten in die Auftragsannahme - konkretisiert für Unternehmen der eisen- und stahlerzeugenden Industrie. *ZFB Special Issue*, (4):125–151.

Stadtler, H. (2008). Production planning and scheduling. In Stadtler, H. and Kilger, C., editors, *Supply Chain Management and Advanced Planning*, pages 199–216. Springer Berlin Heidelberg.

Talluri, K. and Van Ryzin, G. (1998). Analysis Network of Controls for Revenue Management. *Management Science*, 44(11):1577–1593.

Talluri, K. and Van Ryzin, G. (2004). *The theory and practice of revenue management*. Kluwer Academic Publishers.

Talluri, K. T. and van Ryzin, G. J. (1999). Randomized linear programming method for computing network bid prices. *Transportation Science*, 33(2):207–216.

Tang, L., Liu, J., Rong, A., and Yang, Z. (2001). A review of planning and scheduling systems and methods for integrated steel production. *European Journal of Operational Research*, 133(1):1–20.

Tsai, K.-m. and Wang, S.-c. (2009). Multi-site available-to-promise modeling for assemble-to-order manufacturing: An illustration on TFT-LCD manufacturing. *International Journal of Production Economics*, 117(1):174–184.

Volling, T., Eren Akyol, D., Wittek, K., and Spengler, T. S. (2012). A two-stage bid-price control for make-to-order revenue management. *Computers & Operations Research*, 39(5):1021–1032.

Watanapa, B. and Techanitisawad, A. (2005). Simultaneous price and due date settings for multiple customer classes. *European Journal of Operational Research*, 166(2):351–368.

Williamson, E. L. (1992). *Airline network seat inventory control: Methodologies and revenue impacts*. PhD thesis, MIT, Cambridge.

Winands, E. M. M., Adan, I. J. B. F., and Van Houtum, G. J. (2011). The stochastic economic lot scheduling problem: A survey. *European Journal of Operational Research*, 210(1):1–9.

Wismer, D. a. (1972). Solution of the Flowshop-Scheduling Problem with No Intermediate Queues. *Operations Research*, 20:689–697.

Wu, X. and Zhou, X. (2008). Stochastic scheduling to minimize expected maximum lateness. *European Journal of Operational Research*, 190(1):103–115.

Zhao, Z., Ball, M., and Kotake, M. (2005). Optimization-based available-to-promise with multi-stage resource availability. *Annals of Operations Research*, 135(1):65–85.

Curriculum vitae

Hendrik Guhlich

Personal Information

Place of birth	Essen, Germany
Nationality	German

Education

06/2011—04/2016	Doctoral Candidate Business School, University of Mannheim
10/2005—02/2011	Studies in Mathematics (Diploma) Research Institute for Discrete Mathematics, University of Bonn
08/1995—06/2004	Abitur, Gymnasium Essen-Überruhr

Professional Experience

06/2011—05/2016	Research Assistant, Chair of Logistics and Supply Chain Management & Chair of Production Management, University of Mannheim
02/2008—03/2011	Student Assistant, Research Institute for Discrete Mathematics, University of Bonn