Dissertation

Data driven approaches increasing robustness, accuracy, and service levels of industrial demand fulfilment

Alexander M. Seitz

TECHNICAL UNIVERSITY OF MUNICH

Munich 2017

TUM School of Management
Production and Supply Chain Management
Technical University of Munich, DE-80333 Munich, Germany

TUM School of Management

Production and Supply Chain Management

Data driven approaches increasing robustness, accuracy, and service levels of industrial demand fulfilment

Dipl.-Ing. Univ. Alexander M. Seitz, M.Sc.

Vollständiger Abdruck der von der Fakultät für Wirtschaftswissenschaften der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Wirtschaftswissenschaften (Dr. rer. pol.)

genehmigten Dissertation.

Vorsitzende(r): Prof. Dr. Stefan Minner
Prüfer der Dissertation:
1. Prof. Dr. Martin Grunow
2. Prof. Dr. Herbert Meyr

Die Dissertation wurde am 12.01.2017 bei der Technischen Universität München eingereicht und durch die Fakultät für Wirtschaftswissenschaften am 15.04.2017 angenommen.

Abstract
Motivated by the recent advances of information technology and big data tools, this thesis investigates how the robustness and accuracy of demand fulfilment as well as service levels can be increased in industrial settings by systematically exploiting newly available data and setting the right framework conditions in customer contract portfolio management. The studied system consists of customer contract portfolios directly influencing demand fulfilment, the processes demand planning and supply network planning, the demand fulfilment processes customer segmentation, allocation planning, and order promising, as well as the interface of these processes to the set of customers.

After discussing the role of demand fulfilment in supply chain planning and reviewing the relevant literature from several related disciplines, first, a data driven supply chain planning framework is presented, which surpasses current industry practice. The framework streamlines all planning decisions of the system and incorporates demand fulfilment relevant supply chain and customer data. Enabler processes are integrated, which provide the planning processes with data on the capabilities of the supply chain in terms of flexibility, the customer forecasting and ordering behaviour in terms of the accuracy of advance demand information and the length of customer order lead times, and the contractual, strategic, and operational obligations of the supplier towards its customers.

Following, parts of this framework are detailed through the development of three methods that increase demand fulfilment performance by exploiting big data. An order promising method is presented, which anticipates supply changes due to unforeseen demand arrivals in environments with heterogeneous customer order lead times. Product and process flexibilities of supply chains are identified, formalized, and used to represent supply chain flexibility in ATP information. Product flexibility is the possibility to produce several kinds of products from one predecessor product. Process flexibility is the possibility to use one production process to manufacture several products. The method uses shop floor information on individual product and process level. A numerical study based on a case from the semiconductor industry demonstrates that the method increases the accuracy and robustness of order promises.

Additionally, two demand fulfilment approaches considering customer profitability, the accuracy of advance demand information provided by customers, and the lengths of customer order lead times are proposed. The methods exploit data on individual customers and products by allocating supply on a highly granular level at high planning frequencies. Numerical studies show that the approaches support efficient supply allocation, lowering inventory levels, and increase service levels, especially for customers with truthful forecasts. Consequently, the planning security is raised for the supplier because customers are incentivised to provide truthful advance demand information.

On basis of these approaches, industrial contract portfolios with customer-specific terms are analysed in order to derive insights aiding suppliers in their contract portfolio management and in their design of demand fulfilment processes. The analysis shows that demand fulfillment performance is not primarily determined by the absolute length of the order lead times but by the presence of a negative correlation with the accuracy of advance demand information in the entire customer contract portfolio. Consequently, suppliers must consider the portfolio of all customers and negotiate relatively long order lead times for customers showing relatively low accuracy of advance demand information.

Finally, all results, contributions, and limitations of the presented work are discussed with regards to the studied system as a whole. Managerial insights are derived and possible directions for future research are outlined.

Zusammenfassung
Vor dem Hintergrund tiefgreifender Fortschritte in der Verarbeitung großer Datenmengen (Big Data) werden Verfahren entwickelt, die durch systematische Nutzung von Daten und das richtige Setzen von Rahmenbedingungen im Kundenvertragsportfolio-Management die Robustheit und Genauigkeit von Auftragszusagen im Demand Fulfilment, sowie das Kundenservicelevel in industriellen Umfeldern erhöhen. Das untersuchte System besteht aus Kundenvertragsportfolios, welche das Demand Fulfilment direkt beeinflussen, den Planungsprozessen Demand Planning und Supply Network Planning, den Demand Fulfilment-Prozessen Customer Segmentation, Allocation Planning und Order Promising, sowie deren Schnittstellen zu externen Kunden.

Nach der Einordnung des Demand Fulfilment in die Supply Chain-Planung und der Diskussion des Standes der Wissenschaft in relevanten Forschungsgebieten wird zunächst ein datenbasiertes Supply Chain Planungsframework dargestellt, welches die aktuelle industrielle Praxis weiterentwickelt. Das Framework koordiniert die Planungsentscheidungen des betrachteten Systems und integriert Demand Fulfilment-relevante Supply Chain- und Kundendaten. Darüber hinaus beinhaltet das Framework sogenannte Enabler-Prozesse, die den Planungsprozessen Daten bezüglich der Flexibilitäten in der Supply Chain, des Prognose- und Bestellverhaltens von Kunden und der vertraglichen, strategischen und operationellen Verpflichtungen des Zulieferers gegenüber Kunden bereitstellen. Daten bezüglich des Kundenverhaltens beinhalten dabei Informationen zur Genauigkeit von Advance Demand Information und Auftragsvorlaufzeiten.

Anschließend werden drei Methoden für Teile des Frameworks entwickelt, die die Leistungsfähigkeit des Demand Fulfilment durch die Nutzung von großen Datenmengen erhöhen. Es wird eine Order Promising-Methode für industrielle Umfelder mit heterogenen Auftragsvorlaufzeiten entwickelt, welche in der Lage ist, Supply-Veränderungen zu antizipieren, die durch unvorhergesehene Nachfrageschwankungen hervorgerufen werden. Dazu werden vorhandene Flexibilitäten der Supply Chain in ATP-Information abgebildet. Die genutzten Informationen beziehen sich auf Produkt- und Prozessflexibilitäten der Supply Chain, welche identifiziert und formalisiert werden. Produktflexibilität wird als die Möglichkeit definiert, mehrere unterschiedliche Arten von Produkten aus einem Vorgängerprodukt zu erstellen. Prozessflexibilität wird als die Möglichkeit definiert, einen Produktionsprozess für die Herstellung mehrerer unterschiedlicher Produkte zu verwenden. Die Methode nutzt dabei Daten bezüglich individueller Produkte und Prozesse aus der Fertigung. Eine numerische Studie, welche mit Daten aus der Halbleiterindustrie durchgeführt wird, weist die Fähigkeit der entwickelten Methode nach, die Genauigkeit und Robustheit von Auftragszusagen zu erhöhen.

Darüber hinaus werden zwei Demand Fulfilment-Verfahren vorgeschlagen, welche Daten über die Kundenprofitabilität, die Genauigkeit von vom Kunden bereitgestellter Advance Demand Information und die Länge von Kundenauftragslaufzeiten verarbeiten. Die Methoden verwenden Daten individueller Produkte und individueller Kunden und führen die Supplyallokation mit hoher Planungsfrequenz auf sehr feinen Granularitätsstufen durch. Numerische Studien zeigen die Eignung der Verfahren, die Effizienz der Supplyallokation sowie den Servicelevel für Kunden zu erhöhen und gleichzeitig Lagerbestände zu verringern. Durch die besonders ausgeprägte Erhöhung der Servicelevel für Kunden mit hoher Prognosegenauigkeit

werden entsprechende Anreize für Kunden gesetzt, welche die Planungssicherheit für den Lieferanten erhöhen.

Auf Basis dieser Verfahren werden Portfolios Kunden-individueller Verträge analysiert, um Erkenntnisse abzuleiten, welche Lieferanten in der Gestaltung des Kundenvertragsportfolios und des Demand Fulfilment-Prozesses unterstützen. Die Analyse zeigt, dass die Güte des Ergebnisses des Demand Fulfilment Prozesses nicht primär durch die absolute Länge der Kundenauftragsvorlaufzeit bestimmt wird, sondern durch die Existenz einer negativen Korrelation mit der Genauigkeit der Advance Demand Information im Portfolio. Folglich muss beim Verhandeln neuer Verträge das gesamte Vertragsportfolio berücksichtigt werden. Mit Kunden, deren Advance Demand Information eine relativ niedrige Genauigkeit aufweist, müssen lange Auftragsvorlaufzeiten verhandelt werden.

Abschließend werden die Ergebnisse, Beiträge und Beschränkungen der Arbeit ganzheitlich im Lichte des betrachteten Systems diskutiert und mögliche Wege für zukünftige Forschung aufgezeigt.

Acknowledgements

Set up as a collaboration between the group of Production and Supply Chain Management at the Technical University of Munich and the Supply Chain Innovation department of the Infineon Technologies AG, this dissertation was supported by a number of people.

First, I would like express my utmost gratitude to my academic supervisor Prof. Dr. Martin Grunow. His guidance, steady support, and invaluable input did not only challenge and motivate me, but also shaped, sharpened, and developed my thinking.

I am also deeply thankful to my second academic supervisor Prof. Dr. Renzo Akkerman for his opinions, helpful advises, and constructive feedback.

Hans Ehm has been a great manager. The way he leads his employees by understanding and considering their personal situation as well as professional and private goals is both outstanding and inspiring. I am cordially grateful for the opportunities he gave me to develop myself, which will surely continue to help me in my future life.

I further want to thank Dr. Markus Pfannmüller for the countless discussions on demand fulfilment and his feedback regarding this dissertation. Jens Wittstock, Thorsten Domke, Peter Dressler, Dr. Ronald Kometer, and Thomas Polzer supported this project with input especially during its beginning.

Prof. Dr. Herbert Meyr reviewed this thesis with exceptional care and thoroughness. I greatly appreciate his detailed constructive feedback, which improved the final print version substantially.

My colleagues and fellow doctoral candidates have been inspiration and support throughout the entire project; thank you!

Lastly, I wish to express my profound gratitude to my family and friends for their support, motivation, friendship, help, and all the good times that make life worth living. In particular, I am sincerely thankful to my parents for giving me the opportunity and strength to do this dissertation and Anna for always being supportive, patient and loving, for listening and discussing, and for setting a perspective to bring this work to a successful end.

March 2017
Alexander Seitz

Contents

List of figures

List of tables

List of abbreviations

AATP	Allocated available-to-promise
ADI	Advance demand information
APS	Advanced planning system
ATO	Assemble-to-order
ATP	Available-to-promise
A&C	Availabilities and capabilities
BOM	Bill of material
CAP	Conventional allocation planning
CATP	Cumulated available-to-promise
CTP	Capable-to-promise
CUM	CATP type considering ATP, BOM coefficients, and resource consumption
DDAP	Data driven allocation planning
EDI	Electronic data interchange
ERP	Enterprise resource planning
FCFS	First come first served
MTO	Make-to-order
MTS	Make-to-stock
OLT	Order lead time
OTSL	On-time service level
PROC	CATP type considering ATP and resource consumption
PAS	Profitability accuracy score
PROD	CATP type considering ATP and BOM coefficients
SMAPE	Symmetric mean absolute percentage error
SUM	CATP type only considering ATP information
TSL	Total service level

List of symbols

α	Weight factor for customer forecast accuracy.
α^*	Service-level-optimal α.
$\delta a_{g'gg^*}$	Ratio of resource consumption factors of two sequences $S_{g'g}$ and S_{g^*g}.
$\delta N_{g'gg^*}$	Ratio of $N_{g'g}$ and N_{g^*g} of two sequences $S_{g'g}$ and S_{g^*g}.
Θ	Set of combinations θ.
Θ_g	Set of combinations θ containing g as produced product.
Θ'_g	Set of combinations θ containing g as transformed product.
Θ_j	Set of combinations θ containing process j.
θ	Index of existing combinations of resource j, predecessor product g', and successor product g.
ξ^i	Per-unit inventory holding cost.
ξ^p	Per-unit production cost.
$\xi^e_{t\tau}$	Per-unit cost for early fulfilment of demand in period t being due in period τ.
$\xi^l_{t\tau}$	Per-unit cost for late fulfilment of demand in period t being due in period τ.
π	Weight factor for customer profitability.
ρ_d	Base revenue of demand d.
ρ_{dt}	Revenue generated by fulfilling one unit of demand d in period t.
ω	Weight factor for customer order lead time.
a_θ	Resource consumption factor for combination θ.
$aatp_{kt\tau}$	AATP quantity consuming ATP becoming available in period t, which is reserved for demand from segment k being due in period τ.
$aatp_i^{DDAP}$	AATP quantity for customer i resulting from data driven allocation planning.
$aatp_k^{CAP}$	AATP quantity for segment k resulting from conventional allocation planning.
acc_i	Historical accuracy of demand forecasts of customer i.
acc_i^{norm}	Normalised historical accuracy of demand forecasts of customer i.
acc_k^{norm}	Normalised historical accuracy of demand forecasts of segment k.
acc_{ih}^{fc}	Historical accuracy of demand forecasts with horizon h from customer i.
atp	Available ATP quantity (single product, single period case).
atp_t	ATP quantity becoming available in period t (single product case).
atp_{gt}	ATP quantity for product g becoming available in period t.
b_i	Historical forecast bias of customer i.
b_{ih}	Bias of demand forecast provided by customer i with horizon h.
b_k	Historical forecast bias of segment k.
$c_{kt'}$	Consumed portion of AATP for segment k becoming available in period t' used for order promising.
c_t	Consumed portion of CATP becoming available in period t used for order promising.
$catp_t$	CATP quantity becoming available in period t (single product case).
$catp_{gt}^{COM}$	CATP quantity of type COM for product g in period t.
$catp_{gt}^{PROC}$	CATP quantity of type PROC for product g in period t.
$catp_{gt}^{PROD}$	CATP quantity of type PROD for product g in period t.
$catp_{gt}^{SUM}$	CATP quantity of type SUM for product g in period t.

$CT_{g'g}$	Cycle time of sequence $S_{g'g}$.
ct_θ	Production cycle time for combination θ.
D	Set of generic demands (orders and forecasts)
D_g	Set of all generic demands requesting product g.
d	Index for generic demands (order or forecast)
$dist_{ij}$	Distance between customer i and j in terms of their customer score.
dq_{ot}	Delivered quantity for order o in period t.
$\overline{e}_{ih}$	Average error of demand forecast provided by customer i with horizon h.
$e_{it\tau}$	Error of historical demand forecast $d_{it\tau}^{hist}$ given by customer i in period t for period τ.
err_{ih}^{fc}	Historical error of demand forecasts with horizon h from customer i.
$\overline{err}^{acc}$	Average error of customer forecast accuracy scores.
F_g	Set of generic demands d representing demand forecasts for product g.
f_i	Demand forecast of customer i (single period case).
$f_{i\tau}$	Demand forecast of customer i being due in period τ.
$f_{it\tau}^{hist}$	Historical demand forecast given by customer i in period t with due date in period τ.
G_{g^*t}	Set of products g, whose ATP quantities can be used in period t to build CATP for product g^*.
g	Index for generic product (raw material, intermediate product, or finished product).
h	Index for forecast horizon (i.e. time between forecast entry and indicated demand due date)
I	Set of customers.
I_k	Set of customers belonging to segment k.
i	Index for customers.
i^p	Index of most profitable customer.
i^*	Index of ordering customer.
inv_{gt}	Inventory of product g at the end of period t.
inv_g^{-1}	Starting inventory of product g.
J	Set of resources.
j	Index for processes.
K	Set of customer segments.
K_i	Set of segments from which customer i is allowed to consume supply.
k	Index for customer segments.
k^*	Index of segment of ordering customer.
M	Set of intermediate products
$N_{g'g}$	Number or units of g being produced from g' on $S_{g'g}$.
n_θ	Bill of material coefficient for combination θ.
O	Set of customer orders.
O_i	Set of all orders from customer i.
o	Index of customer orders.
olt_i	Average order lead time of customer i.
olt_i^{max}	Maximum order lead time of customer i.

olt_i^{min}	Minimum order lead time of customer i.
olt_i^{norm}	Normalised average order lead time of customer i.
P	Set of finished products.
$p_{it\tau}$	Delivery quantity consuming supply becoming available in period t, which is promised for demand from customer i being due in period τ.
PAS_i	Profitability accuracy score of customer i.
PAS_k^{seg}	Profitability accuracy score of segment k.
$prof_0$	Base profit for promising an order (single product case).
$prof_i$	Per-unit profitability of customer i.
$prof_k$	Per-unit profitability of segment k.
$prof_i^{norm}$	Normalised profitability of customer i.
$prof_k^{norm}$	Normalised profitability of segment k.
$prof_t$	Per-unit profit generated by fulfilling an order in period t (single product case).
$profit_i^{CAP}$	Total profit achieved with conventional allocation planning.
$profit_i^{DDAP}$	Total profit achieved with data driven allocation planning.
$promise_i^{CAP}$	Promise for customer i resulting from conventional allocation planning
$promise_i^{DDAP}$	Promise for customer i resulting from data driven allocation planning.
q_d	Requested quantity of demand d.
q_i	Order quantity of customer i.
q_o	Order quantity of order o (single period case).
$q_{i\tau}$	Demand quantity of customer i in period τ.
$q_{i\tau}^{hist}$	Quantity of historical demand placed by customer i with a due date in period τ.
R	Set of raw materials.
$r(olt_i, acc_i)$	Pearson Correlation of order lead time and forecast accuracy of advance demand information in the entire set of customers.
$RNK()$	Ranking function.
$S_{g'g}$	Sequence of processes transforming product g' into product g.
$score_i^{cust}$	Customer score of customer i.
$score_k^{seg}$	Segment score of segment k.
s_{min}	Minimum size for customer segments.
SL^{CAP}	Service level resulting from conventional allocation planning.
SL^{DDAP}	Service level resulting from data driven allocation planning.
$\overline{stock}$	Average stock level.
$stock_t$	Stock level at the end of period t.
T	Set of time periods.
T^{hist}	Set of historical time periods for calculation of customer ordering behaviour.
T^s	Set of time periods in the total horizon of the experiment.
t, τ	Index for time period.
t_d / t_o	Due date of demand d / order o.
t_o^d	Realised delivery period for order o.
t_o^p	Promised delivery period for order o.
t_o^{placed}	Time period of placement of order o.
u	Upward nesting level.
v_{ik}	Decision variable assigning customer i to segment k.

$\overline{w}$	Maximum distance between any two customers belonging to the same segment.
x_{dt}	Delivery quantity used to fulfil demand d in period t.
$Y_{\theta t}$	Already started, but not yet finished production quantities for combination θ in period t.
$y_{jg'gt}$	Production quantity for production of product g started in period t with predecessor product g' on resource j.
$y_{\theta\tau}$	Production quantity for combination θ in period τ.

1 Introduction

"Next generation global order management leverages the power of technologies including Web 2.0, mobility, big data analytics, social media, and cloud-based solutions."
(TATA Consultancy Services 2013)

1.1 Motivation and background

Viewing their supply chain as a potential competitive advantage, companies raise resource utilization and reduce inventories in order to increase their operational efficiency. At the same time, companies have to focus on their core competencies in order to maintain their competitiveness in global markets. This specialization leads to an increasing number of partners interacting in supply chains and a constant rise of supply chain complexity. Moreover, an increase in merge and acquisition activities and a growing number of demand fulfilment channels, which often result in heterogeneous and misaligned planning systems, further challenge supply chain management.

In consequence, supply chains become increasingly sensitive to human and system-caused disturbances. Simultaneously, such disturbances appear more frequently as economic cycles are shortening and growing supply chain sizes lead to a more pronounced bullwhip effect, i.e. an increase of demand variation and uncertainty when demand moves upstream in a supply chain (see e.g. Lee et al. 2004). The reasons for these distortions are, on the one hand, of behavioural nature. Customers do not share their private information with suppliers, who, in turn, misinterpret demand signals they receive (see e.g. Kilger and Meyr 2015 or Vogel 2014). On the other hand, Lee et al. (2004) list the four operational causes rationing gaming, order batching, forecast updating, and pricing, which increase the amplitude of demand in every echelon of the supply chain.

Oftentimes, the operational flexibilities of companies to react to sudden demand distortions are limited. Long production cycle times force forecast-based production starts long before customers place their orders. Additionally, short and further decreasing product life cycles limit the possibilities to compensate demand uncertainty with increased buffer stocks, because the risk of obsolescence of stocks is too high. Furthermore, capital intensive capacities constrain the possibilities to increase manufacturing flexibility to react to unforeseen demand changes with adapted production volumes.

In consequence, periods of supply shortage, in which the supply output of the chain is maximised and cannot be adapted to demand changes anymore, occur more frequently. Then, the demand fulfilment decision, i.e. when to fulfil which (current or future) customer demand with which supply is of utmost importance. However, the capability of companies to fulfil orders as required, meet contractual obligations towards the customers, keep customer satisfaction on a high level, and maintain a high customer retention rate is challenged by short and heterogeneous customer order lead times (OLTs), low customer forecast accuracy, and frequent order cancellation and rescheduling, all caused by the changing demand fulfilment expectations of customers in demand driven markets.

A typical example for the above described situation is the semiconductor manufacturing industry. The companies of this sector are usually situated in upstream positions of their supply chains. Therefore, they are exposed to severe demand fluctuations due to the bullwhip effect.

For example, the global semiconductor market without memory and microprocessors shrank by almost 40% in 2009 while it grew by over 50% in 2010. In the same timeframe, the global gross domestic product only varied between -4% and +4% (WSTS Inc. 2015). Additionally, customer OLTs are with a maximum of six weeks relatively short compared to the production cycle times, which are up to four times as long. Therefore, production needs to be started based on internal and external forecasts, which, because of their long horizon, are subject to substantial uncertainties. However, because of short product life cycles of only a few months holding high buffer stocks implies a high risk of obsolescence. On the other hand, as semiconductor manufacturing equipment is highly capital intensive, the possibilities to buffer demand fluctuations with additional capacity are limited.

A survey performed by Oracle and Capgemini in 2013 among 589 supply chain executives from the manufacturing, high-tech, and retail industries shows that, under the above described circumstances, companies are struggling to fulfil customer demands on-time and according to initially given commitments. The study reveals that 42% of manufacturing and high-tech companies view accurately promising delivery dates as the main challenge in maintaining customer satisfaction. Additionally, inaccurate order promising is seen as one of the main cost drivers for demand fulfilment as it causes additional efforts such as procuring costly external production resources in the short-term, triggering emergency processes in logistics, or intensifying communication with suppliers and customers in order to meet committed delivery dates (Oracle and Capgemini 2013). Another survey from 2013, performed by the Aberdeen Group among 151 chief supply chain officers shows that even best-in-class companies are unable to forecast their customer demands with more than 85% accuracy on the product family level, leading to the need for expensive buffer stocks and additional capacities in order to compensate this lack of forecast accuracy with increased supply chain flexibility (Aberdeen Group 2013).

All this shows that companies need new demand fulfilment approaches in order to cope with the above mentioned challenges. To exploit the full potential of these approaches, they also need to understand the interactions between demand fulfilment and other processes setting framework conditions for demand fulfilment. The recent advances in big data technology, i.e. the new possibilities of data storage, data mining, data exchange, and data analysis, are seen as enablers for research and practice to develop new demand fulfilment approaches that improve the performance of current state of the art methods (e.g. TATA Consultancy Services 2013). For example, with the newly available data, companies can monitor the forecast accuracy and OLTs, i.e. the time between order placement and requested delivery date, of their customers on a high granularity. This higher transparency of the ordering behaviour of customers provides opportunities to increase the efficiency, accuracy, and robustness of demand fulfilment activities. Moreover, the new technologies also provide transparency of supply chain characteristics and flexibilities that can be exploited in demand fulfilment to significantly increase its performance.

While big data applications are widely studied in, e.g., marketing, sales, finance, product development, compliance, and fraud prevention, only little initial research exists that deals with the exploitation of data in demand fulfilment. This thesis is a first step towards the systematic exploitation of data in industrial demand fulfilment approaches in order to increase its robustness, accuracy, and service levels. It is furthermore the first studying the interrelation of

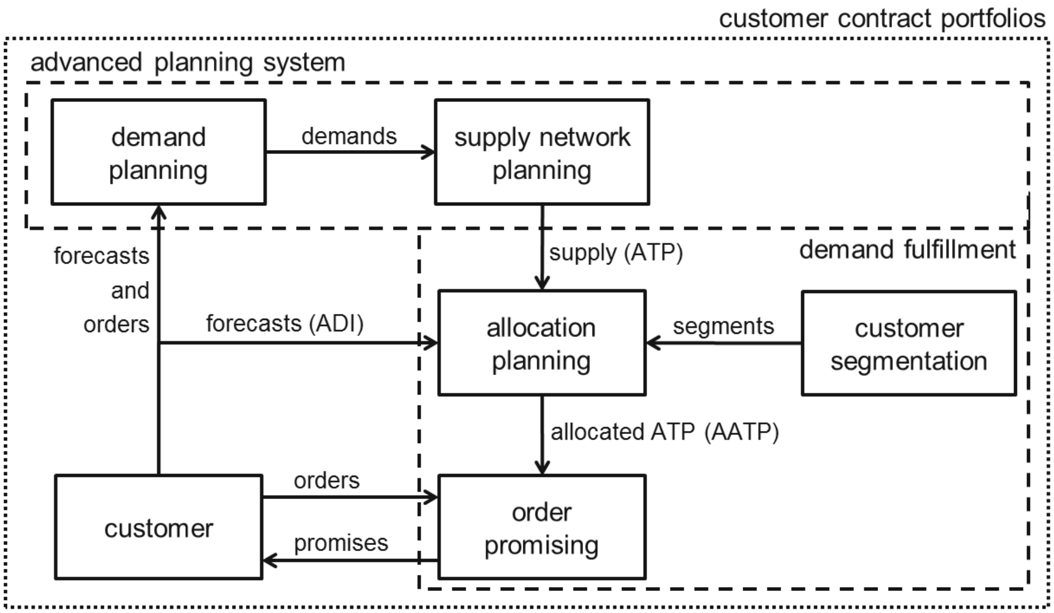

Figure 1: Overview of studied system

the terms in customer contract portfolios and the order promising and supply allocation rules in demand fulfilment.

1.2 Studied system

The software tools supporting demand fulfilment in industrial environments are typically enterprise resource planning (ERP) and advanced planning systems (APS). This thesis focusses on demand fulfilment related processes, which are usually implemented in APS. Figure 1 provides an overview. In the context of APS, typically only the processes customer segmentation, allocation planning, and order promising are seen as demand fulfilment processes. They are usually implemented in a respective software module. Other relevant processes are the demand planning and the supply network planning process, which provide necessary input for the three demand fulfilment processes. Customer contract portfolios set framework conditions for demand fulfilment. In the following, the purposes, objectives, inputs, and outputs of the processes shown in Figure 1 are outlined shortly. More details are provided in Chapter 2.

Dyadic contracts between customers and suppliers set the terms and conditions for the exchange of demand information, i.e. the exchange of customer demand forecasts, also called advance demand information (ADI), and customer orders. The contractual terms include the horizon and maximum volatility of ADI as well as minimum order lead times and maximum allowed deviations between order and ADI. All contracts in place between a supplier and their customers, called the customer contract portfolio, set the framework conditions for the demand fulfilment process of the supplier.

On basis of customer forecasts and orders as well as internal forecasts coming from marketing, sales, and operations functions, demand planning generates short- and mid-term forecasts for the total market demand that is addressed to the supply chain. The forecasts typically include contingencies in form of planned buffer stocks in order to deal with uncertainty. The objective of the process is to forecast total market demands as accurately as possible. The results are provided to the supply network planning process.

In the supply network planning process, the total supply chain demands are matched with the total supply chain resources to generate a global master production schedule. Resources are, amongst others, machine capacities, stocks, work in progress, and raw materials. Often, supply network planning methods only use bottleneck capacity information, i.e. the capacity of machinery, which is constraining production. Other machinery is then modelled as a time delay between the bottlenecks. The objective of the supply network planning process is to generate a master production schedule that utilizes available supply chain resources at a certain target level and fulfils supply chain demands according to their due date. From the master production schedule, the supply information is generated, which is used in the demand fulfilment processes to generate delivery date confirmations to the customers. This supply information is called available-to-promise (ATP) and comprises current inventory as well as planned supply receipts from production.

The customer segmentation process clusters customers into segments based on their importance for demand fulfilment. The objective of the process is to generate segments in which customers are as homogeneous regarding their importance as possible. Typically, the importance of customers is determined by their profitability for the company. Other approaches also consider the strategic importance of customers.

Based on ADI provided by the customers, the allocation planning process generates ATP reservations, called allocated available-to-promise (AATP), for customer segments. The AATP is determined such that the demand of important customers is fulfilled with priority. Note that, besides supply network planning, allocation planning is the second interface of supply chain planning with ADI from the customer.

The AATP is then forwarded to the order promising process which generates delivery date commits for incoming customer orders in real-time, i.e. a first come first served (FCFS) manner. Here, AATP quantities can be segmented or nested. If they are segmented, the order promising process can only exploit AATP in the time dimension, if not enough AATP exists in the planning period of the order due date. If AATP is nested, supply that is reserved for customer segments with lower importance can be consumed additionally.

1.3 Problem statement

Demand fulfilment has been studied extensively in the literature. However, the presented approaches oftentimes ignore important aspects of demand fulfilment in industrial practice or are tailored for very specific problem settings, making it difficult to apply the solutions to other, more general, problem settings. Many challenges and characteristics of industrial settings have not been addressed so far. Some of these are:

- Demand fulfilment processes stand in the context of short-, mid- and long-term demand planning and supply network planning. Allocation planning and order promising must be considered in this context in order to increase the robustness and accuracy of demand fulfilment as well as customer service levels.
- Customer OLTs are heterogeneous leading to hybrid make-to-order (MTO), make-to-stock environments (MTS).

- Production is planned on finished product level whereas demand is forecasted on product family level. Therefore, demands used to plan production are uncertain regarding the mix on finished product level.
- Supply network planning exploits supply chain flexibilities to change the master production schedule when unforeseen demand realises. These changes are not anticipated in current order promising approaches.
- Customers inflate their demand forecasts (i.e. ADI) strategically in order to game the allocation planning procedure of their suppliers.
- Customers provide orders with the minimum OLT contractually agreed upon, which is oftentimes much shorter than production cycle times. Therefore, production must be started on uncertain forecast information.
- Advances in big data tools enable companies to monitor the forecasting and ordering behaviour of their customers on high granularity. Furthermore, data on the status of the shop floor can be provided to, and processed in demand fulfilment processes.
- In supply shortage situations, the accuracy of ADI and the length of OLT interact in the demand fulfilment decision. Especially regarding the length of OLT, customers and suppliers follow conflicting goals. While customers try to negotiate short OLTs in order to reduce uncertainty, suppliers prefer long OLTs, which increase planning security.
- The portfolio of contractual terms and conditions negotiated with customers influence the performance of the demand fulfilment of the supplier. Therefore, suppliers need to consider the interactions between the contracts negotiated with individual customers with regards to overall system performance.

Approaches integrating some of the different planning processes presented in Section 1.2 exist. However, so far no framework has been presented that considers all of these processes and integrates newly available data on high granularity with respect to robust and accurate demand fulfilment in industrial settings. Also, the majority of publications on demand fulfilment presents models for homogeneous customer OLTs and ignores the flexibilities exploited in supply network planning as well as the interdependencies between supply network planning and order promising with regards to the robustness and accuracy of promised delivery dates. Furthermore, the effects of the uncertainty of the realisation of demands on finished product level on these measures have not been studied yet. Some initial steps towards the consideration of heterogeneous customer OLTs and the integration of supply network planning and order promising have been taken. However, until now no order promising approach has been presented that anticipates changes in the master production schedule caused by new order arrivals by integrating shop floor data into the promising decision. Moreover, there is a lack of literature presenting demand fulfilment approaches that increase service levels by considering customer forecasting and ordering behaviour and raise planning security by counteracting the strategic inflation of ADI and incentivizing customers to provide orders with lead times being longer as the contractually agreed minimum. Also, the interactions of contractual agreements of a supplier with its entire customer set with regards to demand fulfilment performance have not been studied so far.

This thesis addresses the following research questions in order to close the gaps described above:

RQ1. How should the supply chain planning processes of the studied system (Figure 1) be integrated in order to increase customer service levels and the robustness and accuracy of demand fulfilment?

RQ2. How should available data on supply chain capabilities be considered in demand fulfilment in order to increase the accuracy and robustness of order promises in industrial settings with heterogeneous customer OLTs and uncertainty regarding the realisation of aggregated demand forecasts on finished product level?

RQ3. How should available data on the forecasting and ordering behaviour of customers be considered in demand fulfilment in order to increase service levels in supply shortage situations?

RQ4. How does integrating data on customer forecasting and ordering behaviour into demand fulfilment processes increase planning security?

RQ5. How should the portfolio of contractual agreements with the entire set of customers be managed in order to maximise the demand fulfilment performance of the supplier?

1.4 Outline of the thesis

This thesis investigates how the robustness and accuracy of demand fulfilment as well as service levels can be increased in industrial settings by exploiting newly available data being enabled by recent advances of information technology and big data tools.

First, the role of demand fulfilment in supply chain planning is discussed in Chapter 2. Then, relevant literature from several related disciplines is reviewed in Chapter 3, which in parts bases on the literature discussions in Seitz et al. (2016a), Seitz and Grunow (2017) and Seitz et al. (2016b). In Chapter 4, a data driven supply chain planning framework for robust and accurate demand fulfilment is presented, which surpasses current industry practice. Chapter 4 is based on Seitz et al. (2016a).

The Chapters 5 to 7 detail parts of this framework and develop methods that exploit big data in order to increase demand fulfilment performance.

In particular, an order promising method that represents supply chain flexibilities in ATP information is developed in Chapter 5. The approach uses shop floor data on individual product and process level. A numerical study shows the superiority of the presented approach compared to conventional order promising and demonstrates increased robustness and accuracy of demand fulfilment. Chapter 5 is based on Seitz and Grunow (2017).

In Chapter 6, an allocation planning approach considering the historical bias of ADI provided by the customers is proposed. The approach allocating supply to individual customers increases service levels and incentivises customers to provide truthful ADI. A numerical case study using data from the semiconductor industry shows that the advantages of the methodology decrease with declining predictive quality of data. Chapter 6 is based on Seitz et al. (2016b).

Chapter 7 extends this approach to also consider data on the OLT of customers. Conclusions on the optimal management of the portfolio of customer contracts are derived from a numerical study testing the developed method for several problem instances with different correlations between the accuracy of ADI and the length of OLT in the set of customers.

Finally, Chapter 8 discusses the results, contributions, and limitations of the research presented in this thesis. Managerial insights are derived from the results and possible directions for future research are outlined.

2 The role of demand fulfilment in supply chain planning

This chapter gives an overview of supply chain planning and discusses the role of demand fulfilment in this context. Also, the nature and sources of available data are presented.

After a generic overview of industrial supply chain planning as a hierarchical planning structure in Section 2.1, the concepts and typical planning approaches for demand planning and supply network planning are illustrated in Section 2.2. Section 2.3 details the demand fulfilment processes customer segmentation, allocation planning, and order promising, presents their interlinkages with supply network planning, and lists typical performance measures. Finally, the sources of data in industrial supply chain planning are discussed in Section 2.4. For a more comprehensive discussion of supply chain management and planning, the interested reader is referred to Vogel (2014).

2.1 Overview of supply chain planning

A supply chain is a network of organizations, whose processes and activities are interlinked by material, information, and financial flows with the purpose to fulfil demands of end customers (see e.g. Stadtler et al. 2015). Industrial supply chains are customer and profit oriented. Their core activities are the procurement of raw materials, the production of semi-finished and finished products, the distribution of these products within the supply chain, and, ultimately, selling the products to customers (Lee and Billington 1993).

The coordination of the activities and flows within a supply chain as well as the management of the relationships of the supply chain partners is the task of supply chain management (see e.g. Christopher 1998). Supply chain management aims at improving the competitiveness of the entire chain through superior – and profitable – customer service. Means to increase competitiveness are, for example, to reduce supply chain costs and cycle times, to increase supply chain flexibility and efficiency, to raise customer service levels, and to improve the robustness and accuracy of supply chain planning (see e.g. Stadtler et al. 2015).

The coordination of these partly conflicting goals requires sophisticated strategic, tactical, and operational planning and decision making. It is a complex planning task that can neither be tackled simultaneously nor sequentially. For simultaneous planning, the problem size simply is too large. Sequential planning, e.g. using the MRP (Orlicky 1975) or MRP II (Wight 1984) concepts, does not consider bilateral interdependencies of the planning tasks adequately (Fleischmann et al. 2015).

Therefore, industrial supply chain planning typically employs the concept of hierarchical planning. Building on the seven principles decomposition, aggregation, coordination, model building, anticipation, disaggregation, and model solving, the hierarchical planning concept divides a decision problem into hierarchically structured planning levels and links these levels with each other so that a feasible solution of good quality for the original problem results (see e.g. Stadtler et al. 2012). Aggregation along the dimensions time and entity (e.g. product, customer, site, etc.) is used to reduce the complexity of the addressed planning problems. Rolling horizon schedules, i.e. the typical iterative re-planning scheme of industrial supply chain planning, result from aggregation along the time dimension. The decisions on the different levels of the hierarchy are coordinated by top-down instructions and bottom-up feedbacks. For each level, models of the decision problem are built. These anticipate the capabilities and potential reactions of the models on subordinate levels and disaggregate the instructions from higher planning levels. By solving the models of the hierarchy in an appropriate manner, the

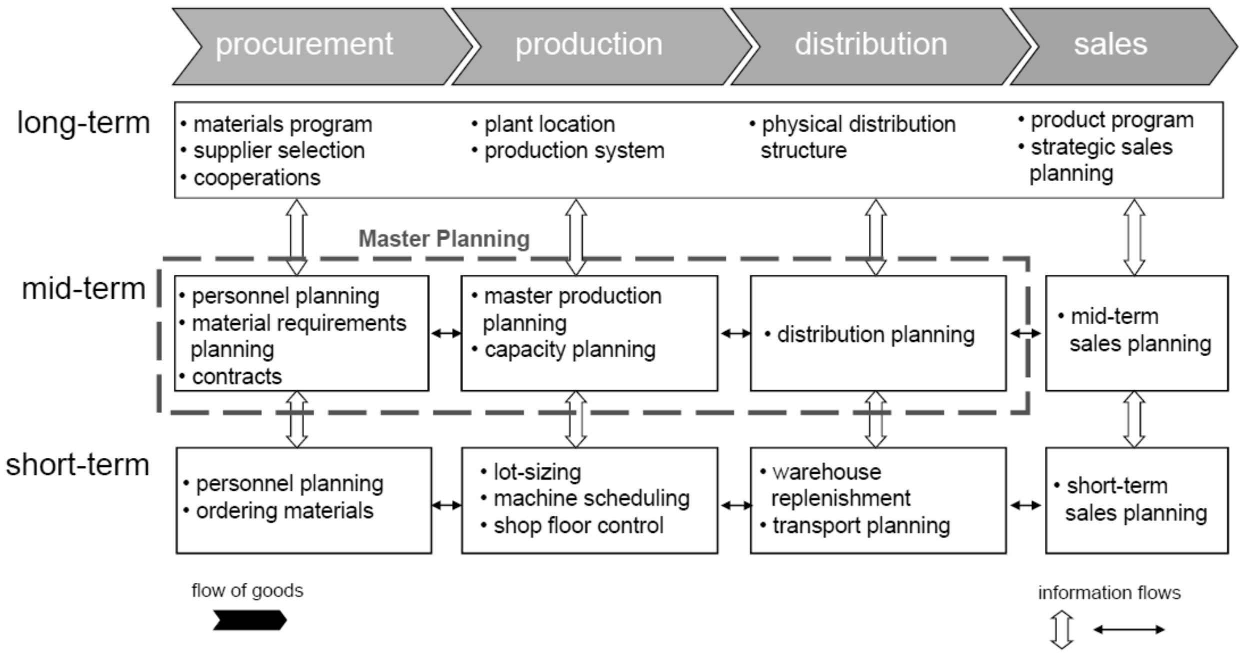

Figure 2: Supply chain planning matrix (taken from Stadtler et al. 2012)

original planning problem is solved. The recent advances in data storage, data mining, data exchange, and data analysis technology enable manifold ways to improve hierarchical supply chain planning in the industry.

For a more detailed discussion of the hierarchical planning concept, see e.g. Fleischmann and Meyr (2003) or Stadtler et al. (2012). Rohde et al. (2000) apply this concept to supply chain planning and complement the hierarchical structure with a functional layer that is oriented at the four above mentioned core activities of supply chains, namely procurement, production, distribution, and sales. The resulting supply chain planning matrix is a framework depicting the interdependencies of supply chain planning tasks in the two dimensions activity and time. Amongst others, the publications of Schneeweiß (2003) and Fleischmann et al. (2015) are developing this concept further. Figure 2 shows a typical representation.

In the time dimension, the matrix differentiates between long-, mid-, and short-term planning tasks, which correspond to strategic, tactical, and operational supply chain management decisions. The supply chain planning problem is decomposed into individual planning tasks, which are interlinked by horizontal and vertical information flows. Vertical interactions represent the instructions and feedbacks from hierarchical planning. Horizontal information flows coordinate the planning of closely interrelated supply chain activities.

The so-called customer order decoupling point is a central characteristic of a supply chain that influences the nature of the planning tasks in the supply chain planning matrix (see e.g. Fleischmann and Meyr 2003). It separates the forecast-driven planning tasks focussing on inventory management from the order-driven planning tasks concentrating on accurate and robust fulfilment of customer orders. Hence, the customer order decoupling point corresponds with the horizon, over which customer orders are known to the supply chain planning processes, i.e. the time between the placement of orders and their due dates. In this thesis, this horizon is called customer OLT. Oftentimes, supply chain planning literature refers to the location of the customer order decoupling point as a managerial decision influencing the time needed to fulfil an order and the value of stocking units that need to be held at this point (see e.g. Stadtler et al. 2012). While this is true for many business-to-customer environments, in industrial business-to-business settings, especially for suppliers located upstream in industrial supply chains, customer OLTs and, consequently, the customer order decoupling point is a given input parameter for supply chain management and planning. This is because customer OLTs are negotiated between suppliers and customers in bilateral contractual agreements. Thereby, customers usually try to negotiate short OLTs in order to shift the supply risk to the suppliers, while suppliers try to set long OLTs in order to reduce planning uncertainties. Depending on the market power of suppliers and customers, customer OLTs are set to different lengths. Moreover, customer OLTs also depend on the specific planning tools and processes employed by different customers. Therefore, customer OLTs in industrial environments are usually heterogeneous.

This thesis studies the system of supply chain planning processes depicted in Figure 1, which are usually implemented in APS. The supply network planning process, which is termed master planning in Figure 2, corresponds to mid-term procurement, production, and distribution planning tasks. The processes demand planning, customer segmentation, allocation planning, and order promising correspond to the mid- and short-term sales planning tasks in the supply chain planning matrix. They provide all other planning tasks with demand information in form of already known and projected future customer demand. Since, as described above, the ultimate goal of supply chain management is to increase competitiveness through superior customer service, i.e. to meet customer demands in the best possible way, these processes are of central importance to supply chain planning. Hence, accurate and robust planning of the corresponding activities is crucial for every supply chain in order to ensure long-term business success.

2.2 Demand planning and supply network planning

2.2.1 Demand planning

The first step towards meeting the ultimate goal of supply chain management, i.e. to increase customer service levels, is to anticipate future customer demands as accurately as possible. Then, subsequent planning activities are able to prepare the provision of supply in a way that customer orders can be fulfilled as close to their requested delivery date as possible while the resources of the supply chain are utilized efficiently (see e.g. Chen et al. 2007). Therefore, demand planning aims at predicting future customer demand as accurately as possible. Furthermore, safety stock levels are planned to buffer uncertainties in this forecast (see. e.g. Kilger and Wagner 2015). Accurately planned demand enables smooth production and procurement planning, because sudden changes in production and procurement plans due to unforeseen lacks or peaks of demand are prevented. This, in turn, leads to reduced cost for inventory and idle capacity (Vogel 2014).

There are three types of techniques to generate the demand forecast: statistical, judgemental, and collaborative forecasting (Kilger and Wagner 2015). Statistical techniques are solely using historical data. The data is analysed by time-series based or causal approaches in order to predict the future development of demands. Judgemental forecasting includes additional information, e.g. on future one-time events, into the forecasting process in order to derive a more accurate forecast. Collaborative forecasting extends the judgemental forecasting approach by involving several internal or external partners into the forecasting process and combining their forecasts and views into one aligned demand forecast. For a deeper discussion of these techniques, the reader is referred to Hanke and Wichern (2008), Makridakis et al. (1998), and Tempelmeier (2008).

The accuracy of demand forecasts is measured by quantifying the forecast error after demand realisation. This is determined by evaluating the difference between the volumes of the forecasted and the realised demands. When analysed over a longer time period, the accuracy of a forecast also quantifies the uncertainty contained in it.

A multitude of measures exists to measure the forecast error. Some of them are the mean error, the mean absolute deviation, the mean squared error, the root mean squared error, and the mean absolute percentage error (Meyr 2012). For a broad overview of measures and a discussion of the advantages and disadvantages, the interested reader is referred to Kilger and Wagner (2015).

According to Fildes and Kingsman (2011) the forecast error is composed of a random and a systematic component. The latter is called bias and refers to a constant over- or underestimation of future demands, which can result from rationing gaming (see Lee et al. 2004).

Although there are many other purposes for which the results of demand planning are used (e.g. financial forecasting, accounting, marketing, or logistic network planning), this thesis focusses on the interactions of demand planning with the supply network planning process. Taking already known future and historical customer orders as well as ADI from customers into account, demand forecasts are generated on aggregated levels, like product family per customer segment. Supply network planning, however, oftentimes matches available resources with demands on finished product (also called stock keeping unit) level (Mentzer and Bienstock 1998). Consequently, the demand forecasts have to be disaggregated for supply network planning purposes. In most industrial settings, this disaggregation is done according to the historical share of individual finished products sold in one product family as well as the current share of products in ADI and already received, but not yet delivered orders.

The disaggregation of demand forecasts causes a second type of uncertainty, which in this thesis is called demand mix uncertainty. It is defined as the uncertainty of the demand forecast with regards to the ratio of the individual product volumes, when the total demand of the product family is given.

In Chapters 4 and 5, the symmetric mean absolute percentage error (SMAPE) (see e.g. Armstrong 1985 or Ott et al. 2013) is used to determine the error of a forecast. It is chosen since it is widely used in the industry. The main reason for its use is its ease of implementation. It does not result in infeasibilities if there is no realised demand but a forecast or no forecast but final demand in certain time periods.

2.2.2 Supply network planning

In order to synchronize and coordinate the flow of materials between suppliers, production sites, warehouses, and customers, supply network planning balances demand with supply chain resources (see e.g. Albrecht et al. 2015). The process usually considers procurement, manufacturing, and deployment lead times and resource capacity constraints. It trades off the cost for additional in-house or external capacity, capacity usage, stock holding, transportation between sites, and late or no fulfilment of demand for the entire supply chain. Supply network planning hence comprises the planning activities, which are called master planning in Figure 2.

In industrial settings, the mathematical models used for supply network planning are usually deterministic and aim at cost minimisation or profit maximisation (Stadtler 2012). Because of the heterogeneity of customer OLTs in industrial practice, companies usually plan their production to the finished product storage location. Otherwise, if production was planned only to an intermediate product storage location, orders coming in with short lead times could not be fulfilled on time.

For a basic linear programming model, the reader is referred to Fleischmann and Meyr (2003). In the following, a supply network planning model is presented, which is based on Leachman (1993). The model, which is used in Chapter 5, decides on the production quantities $y_{jg'gt}$ for production of product $g \in M \cup P$ started in period $t \in T$ with predecessor product $g' \in R \cup M$ on resource $j \in J$. The sets R, M, and P are defined as the sets of raw materials, intermediate products, and finished products, respectively. For better readability $\theta \in \Theta$ is defined as an existing combination of j, g', and g. The combinations θ are characterized by a resource consumption factor a_θ, a bill of material (BOM) coefficient n_θ, and cycle time ct_θ. The BOM coefficient indicates the number of units of products g that are produced from on unit of product g' on resource j. The subsets Θ_j, Θ_g, and Θ'_g are defined as the sets of combinations θ containing process j, g as produced product, and g as transformed product, respectively. The model further decides on the delivery quantities x_{dt} used to fulfil demand $d \in D$ in period t. Because of the heterogeneous customer OLTs in industrial environments, the set D consists of both customer orders and demand forecasts. Equations (1) to (8) describe the production planning model.

Maximise

$$z = \sum_d \sum_t \rho_{dt} x_{dt} - \xi^p \sum_\theta \sum_t y_{\theta t} - \xi^i \sum_g \sum_t inv_{gt}, \tag{1}$$

subject to

$$\sum_t x_{dt} \leq q_d, \qquad \forall d \in D; \tag{2}$$

$$inv_g^{-1} + \sum_{\tau \leq t} \sum_{\theta \in \Theta_g} n_\theta \cdot y_{\theta \tau - ct_\theta} = inv_{gt} + \sum_{\tau \leq t} \sum_{d \in D_g} x_{d\tau}, \qquad \forall g \in P, t \in T; \tag{3}$$

$$inv_g^{-1} + \sum_{\tau \leq t} \sum_{\theta \in \Theta_g} n_\theta \cdot y_{\theta \tau - ct_\theta} = inv_{gt} + \sum_{\tau \leq t} \sum_{\theta \in \Theta'_g} y_{\theta\tau}, \qquad \forall g \in M, t \in T; \tag{4}$$

$$\sum_{\theta \in \Theta_j} a_\theta \cdot y_{\theta t} \leq c_{jt}, \qquad \forall j \in J, t \in T; \tag{5}$$

$$y_{\theta t} = Y_{\theta t}, \qquad \forall \theta \in \Theta, t \in T^{hist}; \tag{6}$$

$$y_{\theta t} \geq 0, \qquad \forall \theta \in \Theta, t \in T; \tag{7}$$

$$x_{dt} \geq 0, \qquad \forall d \in D; t \in T. \tag{8}$$

The objective function (1) maximises profits calculated from the per-unit inventory holding cost ξ^i, the per-unit production cost ξ^p, and the per-unit revenues ρ_{dt} generated from fulfilling one unit of demand d in period t. Here, inv_{gt} is the inventory of product g held at the end of period t. Constraints (2) ensure that the delivered quantity does not exceed the demand quantity q_d. Constraints (3) and (4) are inventory balance constraints with the starting inventory inv_g^{-1} of product g and the set D_g of all demands requesting product g. Constraints (5) are capacity constraints. Constraints (6), in which T^{hist} contains all time periods in the past, fix $y_{\theta t}$ to $Y_{\theta t}$, which is the production quantity started on combination θ in period $t \in T^{hist}$ not yet being finished. Constraints (7) and (8) are non-negativity constraints.

The per-unit revenues ρ_{dt} are determined by Equations (9) and (10), which make sure that on-time fulfilment of demands is most preferable and early fulfilment is preferred over late fulfilment. The parameters ρ_d, q_d, and t_d are defined as a base revenue, the demand quantity, and the due date of d, respectively.

$$\rho_{dt} = \frac{\rho_d}{q_d}(|T| - t_d + t), \qquad \forall d \in D, t \in T | t \leq t_d. \tag{9}$$
$$\rho_{dt} = \frac{\rho_d}{q_d}(|T| - t) \qquad \forall d \in D, t \in T | t > t_d. \tag{10}$$

The result of the supply network planning process is the master production schedule which contains the product quantities to be produced or procured, which must be made available at a certain storage location at a certain point in time. The master production schedule serves as the main input for the subsequent production planning and demand fulfilment processes (Vogel 2014). For demand fulfilment, the information contained in the master production schedule is translated into the so-called ATP information, which defines the quantities of current and future product supply that can be used to fulfil incoming customer orders. For example, the ATP quantities atp_{gt} for product g becoming available in time period t can be calculated using Equations (11), in which F_g is defined as the subset of D containing only demand forecasts for product g.

$$atp_{gt} = \sum_{d \in F_g} x_{dt}, \qquad \forall g \in \mathrm{P}, t \in T. \tag{11}$$

For a discussion of other types of ATP and their generation from the master production schedule, the reader is referred to Geier (2014).

As mentioned above, supply network planning in industrial environments oftentimes matches capacities with demand on finished product level. Therefore, aggregated demand information on product family level, which is provided by demand planning, needs to be disaggregated, leading to uncertainty of demand regarding the product mix. Obviously, this uncertainty is carried on in the ATP information generated from the master production schedule. If customer orders realise that deviate from the demand forecast, the following supply network planning run exploits flexibilities in the supply chain to change the master production schedule and meet requested delivery dates of the customers. Order promises to the customers are typically updated according to the result of supply network planning. Hence, demand mix uncertainty endangers the robustness of real-time order promises, which are generated based on ATP information.

2.3 Demand fulfilment processes and performance measures

Demand fulfilment is defined as the process of handling a customer order after it entered the planning system of a company (Fleischmann and Meyr 2004). Hence, demand fulfilment processes have the most immediate impact on company profits, customer service levels, and, consequently, customer satisfaction and retention. This thesis focusses on the planning system described in Section 1.2, which is typical for industrial suppliers. Note that there are other types of demand fulfilment systems. An overview is given in Pibernik (2005).

Especially in supply shortage situations, the decision when to fulfil which customer demand with what supply is of utmost importance for a company. In industrial settings, this decision must be taken in real-time upon order arrival by an order promising process. Thereby, also uncertain future arrivals of customer orders need to be considered. This is done by a supply allocation planning process that reserves supply for orders from certain customer segments or individual customers based on demand forecasts provided by the customer, i.e. ADI. The customer segments are determined by a customer segmentation process (see e.g. Meyr 2009).

Note that, in supply shortage situations, the relevant interface of ADI with the planning processes of a company changes from supply network planning to allocation planning. This is because the supply output of the chain is maximised and cannot be adapted to demand changes anymore. Consequently, companies must optimise the allocation of given supply to customers on basis of ADI in order to keep service levels on a high level and reduce stock levels that result from excess allocations.

Obviously, the accuracy of ADI strongly influences the efficiency of supply allocations and, thus, the customer service and stock levels. To increase planning security, industrial suppliers therefore try to incentivise their customers to provide their ADI with high accuracy and their orders long in advance of their due dates. Other objectives of demand fulfilment processes are to increase the reliability of real-time order promises, to improve the on-time delivery for customer orders, to raise the number of satisfied orders, and to increase revenue and profitability (see e.g. Kilger and Meyr 2015).

In the following, the processes customer segmentation, supply allocation, and order promising as well as industry-typical performance measures for demand fulfilment are discussed.

2.3.1 Customer segmentation

The idea of customer segmentation is to increase profits, revenues, and customer service levels by exploiting the heterogeneity of customers regarding certain characteristics that determine the importance of customers for demand fulfilment. In industrial environments, customers are typically segmented according to their geographical location. This is because demand planning commonly uses such a segmentation of customers to generate aggregate demand forecasts (see e.g. Kilger and Meyr 2015) and the segmentation is simply taken over for the demand fulfilment processes. However, this type of customer segmentation does not necessarily lead to customer segments that facilitate an improvement of the demand fulfilment performance by exploiting customer heterogeneity regarding characteristics that influence profits, revenues, or service levels (see Meyr 2009).

Meyr (2008) presents a number of methods to determine customer segments for demand fulfilment purposes. Here, another approach that determines customer segments $k \in K$ based on customer scores $score_i^{cust}$ is described. The scores define the importance of a customer $i \in I$ for demand fulfilment. The mixed integer linear program, which is used in Chapter 7, models a K-clustering approach. It uses the distance $dist_{ij}$ between two customers i and j, which is defined as the absolute value of the difference of the customer scores (Equation (12)).

$$dist_{ij} = \left| score_i^{cust} - score_j^{cust} \right| \tag{12}$$

The model is described by Equations (13) to (18). With the binary decision variable v_{ik}, customer i is assigned to segment k. The decision variable $\bar{w}$, called width, is the maximum distance between any two customers i and j belonging to the same customer segment.

Minimise

$$z = \bar{w}, \tag{13}$$

subject to

$$\sum_{k \in K} v_{ik} = 1, \qquad \forall i \in I; \tag{14}$$

$$\bar{w} \geq dist_{ij}\left(v_{ik} + v_{jk} - 1\right), \qquad \forall k \in K; i, j \in I: i < j; \tag{15}$$

$$\sum_{i \in I} v_{ik} \geq s_{min}, \qquad \forall k \in K; \tag{16}$$

$$v_{ik} \in \{0,1\}, \qquad \forall k \in K, i \in I; \tag{17}$$

$$\bar{w} \geq 0, \qquad \forall k \in K. \tag{18}$$

The objective function (13) minimises $\bar{w}$, ensuring that the customer segment with the largest maximum $dist_{ij}$-value is as homogeneous regarding the customer scores $score_i^{cust}$ as possible. Constraints (14) ensure that each customer is assigned to exactly one segment. Constraints (15) state that $\bar{w}$ must be greater or equal to the distance between any two customers i and j belonging to the same segment. Constraints (16) set a minimum segment size s_{min}. Constraints (17) and (18) define v_{ik} as binary and $\bar{w}$ as non-negative, respectively.

A minimum segment size is defined because solving pure K-clustering problems can result in customer segments of very different sizes. Industrial companies, however, usually aim at levelling out the sizes of the customer segments used for allocation planning.

2.3.2 Allocation planning

In order to ensure satisfactory service levels for important customers in supply shortage situations, allocation planning reserves ATP for certain customer segments. Typical allocation planning methods allocate supply based on heuristic rules (see e.g. Kilger and Meyr 2015). The supply reservations, called AATP, are subsequently provided to the order promising process.

In the following, a linear programming based allocation planning model is presented, which is based on Meyr (2009). The model, which is described by Equations (19) to (22), allocates the ATP atp_t becoming available at the beginning of planning time period t to the customer segments k. It is used in Chapter 6 and 7. The allocation is done according to a segment score $score_k^{seg}$ that determines the priority of fulfilling the demand forecast $\sum_{i \in I_k} f_{i\tau}$ from segment k in time period τ using ATP becoming available in t. Here, I_k is defined as the set of customers i belonging to segment k. The AATP quantities $aatp_{kt\tau}$result.

Maximise

$$z = \sum_k \sum_\tau \left[\sum_t \left(score_k^{seg} \cdot aatp_{kt\tau}\right) - \sum_{t \leq \tau}(\xi_{t\tau}^{e} \cdot aatp_{kt\tau}) - \sum_{t > \tau}(\xi_{t\tau}^{l} \cdot aatp_{kt\tau})\right], \tag{19}$$

subject to

$$\sum_t aatp_{kt\tau} \leq \sum_{i \in I_k} f_{i\tau}, \qquad \forall k \in K, \tau \in T; \tag{20}$$

$$\sum_k \sum_\tau aatp_{kt\tau} = atp_t, \qquad \forall t \in T. \tag{21}$$

$$aatp_{kt\tau} \geq 0, \qquad \forall k \in K, t \in T, \tau \in T; \tag{22}$$

The objective function (19) maximises the segment-score-weighted supply allocations and penalises early and late demand fulfilment with the factors $\xi^{e}_{t\tau}$ and $\xi^{l}_{t\tau}$. It ensures that demands of segments with high $score^{seg}_{k}$ values are satisfied with priority. Constraints (20) ensure that the generated AATP quantities do not exceed ADI provided by the customer. Constraints (21) state that the sum of allocated supply quantities must equal the total available ATP quantities. Constraints (22) state the non-negativity of the decision variables.

The accuracy of the demand forecasts that are used to generate the AATP is of high importance for the efficiency of the allocation (Vogel 2014). In industrial practice, this efficiency is endangered by the so-called rationing gaming behaviour of customers (see e.g. Lee et al. 2004). In supply shortage situations, customers often deliberately provide inaccurate, i.e. falsely inflated, demand forecasts that do not reflect their true needs. Consequently, stocks are created for forecasts of highly important customers, which are not consumed by subsequent orders. This results in high storage costs. More importantly, due to the limited supply also the service levels for other customers are reduced.

Typically, allocation planning is viewed as a mid-term planning task (see e.g. Ball et al. 2004). However, in industrial settings the communication with the customer and the exchange of demand data is often fully automated and occurs at high frequencies. Therefore, it is beneficial to run allocation planning as a short-term planning process in order to account for demand forecast changes as soon as possible.

2.3.3 Order promising

Order promising is subdivided into the planning tasks order acceptance and due date setting (see e.g. Framinan and Leisten 2010). The order acceptance decision answers the question whether an order should be delivered or declined. It consists of order reception, e.g. by phone, mail, or online (see e.g. Croxton 2003) and an ATP availability check, which searches for AATP supply that can be consumed to satisfy the order. Typical search dimensions are time, product, customer segment, and geography (see e.g. Meyr 2009 or Kilger and Meyr 2015). If these dimensions are applied, supply can be searched in time periods being earlier or later than the requested delivery date of the order, on substitute products for the requested product, in AATP quantities reserved for other customers, and in different locations (e.g. different distribution centres of the supply chain), respectively. Note that the search for AATP across customers is called nesting.

Due date setting determines a delivery date for the incoming order. In most industrial settings orders do not have to be delivered at the requested delivery date from the customer, but there are delivery windows within which the supplier is free to confirm delivery dates. These are usually negotiated between customers and suppliers. In most cases, due date setting simply consists of confirming the incoming order according to the result of the ATP availability check. For this reason, most publications do not distinguish between order acceptance and due date setting.

Fleischmann and Meyr (2004) additionally mention shortage planning as a part of the order promising problem. This planning task exploits options to confirm orders, if supply of the requested product is scarce in the requested time and location. Shortage planning can also be seen as part of the ATP availability check.

In the following, an order promising model is shown, which is based on Meyr (2009). The linear program, which is described by Equations (23) to (26), decides on the portions of allocated supply $c_{kt'}$, which become available in period t' and are used to fulfil an order of $q_{i^*\tau'}$ product units from customer i^* being due in period τ'. It is used in modified versions in Chapters 5, 6 and 7.

Maximise

$$z = \sum_{k\in K_{i^*}}\left[\sum_{t'}\left(score_k^{seg} \cdot c_{kt'}\right) - \sum_{t'\le\tau'}\left(\xi^e_{t'\tau'} \cdot c_{kt'}\right) - \sum_{t'>\tau'}\left(\xi^l_{t'\tau'} \cdot c_{kt'}\right)\right], \quad (23)$$

subject to

$$\sum_{k\in K_{i^*}}\sum_{t'} c_{kt'} \le q_{i^*\tau'}; \quad (24)$$

$$c_{kt'} \le \sum_{\tau} aatp_{kt'\tau}, \qquad \forall k \in K_{i^*}, t' \in T; \quad (25)$$

$$c_{kt'} \ge 0, \qquad \forall k \in K_{i^*}, t' \in T. \quad (26)$$

The objective function (23) maximises the segment-score-weighted supply consumptions and penalises early and late demand fulfilment with the factors $\xi^e_{t'\tau'}$ and $\xi^l_{t'\tau'}$. Constraints (24) state that the sum of consumed supply must not exceed the ordered quantity. Constraints (25) ensure that the allocation quantities $aatp_{kt'\tau}$ are not exceeded and Constraints (26) define the non-negativity of the decision variables.

The model allows nesting of customer segments. The segments, from which customer i^* is allowed to consume allocated supply, are represented in the set K_{i^*}. This set contains all segments, for which Inequality (27) holds, in which k^* represents the customer segment of the ordering customer.

$$score_k^{seg} \le score_{k^*}^{seg} \quad (27)$$

In industrial environments, early fulfilment of orders is usually not allowed. Therefore, the order promises $p_{i^*t'\tau'}$, i.e. the quantities promised for delivery in period t', are calculated with Equations (28) to (30). Equations (28) state that no shipment takes place before the order due period τ'. Equation (29) defines the promised delivery in the due period τ' as the sum of the supply portions $c_{kt'}$ consuming allocated supply in all periods t' earlier or equal to τ'. Equations (30) state that the promised delivery quantity in all periods t' equals the supply portions $c_{kt'}$ consuming allocated supply in the same period.

$$p_{i^*t'\tau'} = 0, \qquad \forall t' \in T | t' < \tau'. \quad (28)$$

$$p_{i^*t'\tau'} = \sum_{k\in K_{i^*}}\sum_{t\in T|t\le\tau'} c_{kt}, \qquad t' = \tau'; \quad (29)$$

$$p_{i^*t'\tau'} = \sum_{k\in K_{i^*}} c_{kt'}, \qquad \forall t' \in T | t' > \tau'. \quad (30)$$

2.3.4 Performance measures for demand fulfilment

As mentioned in Section 2.3, the objectives of demand fulfilment are to increase the number of satisfied orders, the amount of orders delivered on time, the reliability of real-time order promises, the efficiency of supply allocation, and the profitability of the company.

In this thesis, the number of satisfied orders is measured in terms of total service level (TSL) as defined in Equation (31). It is the total quantity of confirmed deliveries divided by the total quantity of incoming orders.

$$TSL = \frac{\sum_{i \in I} \sum_{\tau \in T} \sum_{t \in T} p_{it\tau}}{\sum_{i \in I} \sum_{\tau \in T} q_{i\tau}} \tag{31}$$

The amount of orders delivered on time is measured by the on-time service level (OTSL) as defined in Equation (32). It is the total quantity of deliveries confirmed according to the requested delivery date of the customer divided by the total quantity of incoming orders.

$$OTSL = \frac{\sum_{i \in I} \sum_{\tau \in T} p_{i\tau\tau}}{\sum_{i \in I} \sum_{\tau \in T} q_{i\tau}} \tag{32}$$

The reliability of real-time order promises is determined by the measures robustness and accuracy, which are defined in Equations (33) and (34), respectively. Robustness is defined as the share of ordered quantities, which are delivered according to their original confirmed date that is determined by real-time order promising. Here, O is the set of all customer orders o that realised over the horizon T. The promised and realised delivery period of an order are denoted with t_o^p and t_o^d. The parameter q_o is the order's requested quantity, and dq_{ot} is the quantity delivered to fulfil order o in period t. Accuracy is defined as the complement of the share of order quantities, initially promised later than finally delivered.

$$robustness = \frac{\sum_{o \in O, t \in T | t_o^p = t_o^d} dq_{ot}}{\sum_{o \in O} q_o} \tag{33}$$

$$accuracy = 1 - \frac{\sum_{o \in o, t \in T | t_o^p > t_o^d} dq_{ot}}{\sum_{o \in O} q_o} \tag{34}$$

As explained above, if supply is allocated efficiently to customer segments, less excess stocks occur from over-estimating demands of high priority customers. Therefore, the efficiency of supply allocation can be measured by the average stock level at the end of every planning period $\overline{stock}$, which is defined by Equation (35). Here, $stock_t$ denotes the stock level at the end of planning period t.

$$\overline{stock} = \frac{\sum_{t \in T} stock_t}{|T|} \tag{35}$$

The profitability of a company is measured by the total profit generated by promising orders as defined in Equation (36), in which $prof_i$ is the per-unit profitability of customer i.

$$total\ profit = \sum_i \sum_\tau \sum_t prof_i \cdot p_{i\tau t} \tag{36}$$

2.4 Data sources

In Section 1.1, the potential benefits of exploiting newly available data enabled by the growing usage of big data tools in industrial settings are illustrated. This section discusses the sources of such data.

As mentioned in Section 2.2.1, the forecasting of future demand is based on external and internal data sources, which are analysed by statistical and judgemental techniques. Statistical forecasting uses historical demand data from internal data sources. By means of time-series analyses, historical customer demand data derived from the database backbones of the ERP system of the company is analysed to predict future demands.

In judgemental forecasting, additional sources of information are exploited in order to improve the accuracy of statistical methods. External data sources are e.g. predictions of the economic cycle of public or private research institutes, press releases of competitors and companies in related industries, and political and economic coverage in media. Internally, classical sources like contractual obligations or reporting from operations, sales, and marketing, but also novel approaches like crowd opinion or cloud data analysis are used. Crowd opinion techniques use the so-called wisdom of the crowd (Surowiecki 2005) by aggregating data coming from a large number of individuals and deriving decisions from the aggregation. The crowd wisdom type cognition, i.e. thinking and information processing, is of particular importance for demand planning since it can be used to predict global market developments. Crowd opinion methods can be fed with cloud data, which is collected anonymously from the internet utilizing user traffic data on webpages. Often large amounts of dynamic and unstructured data are generated, which need to be handled with newly available big data tools.

Such tools are used to statistically analyse and aggregate the collected data to derive decisions from it. For demand planning and demand fulfilment, the analysis of ADI and order data with big data applications is of particular importance, because conclusions on the customer ordering behaviour, i.e. the accuracy of ADI and the length of OLTs of customers, can be drawn. This information can be used in demand planning and allocation planning to evaluate the demand information received from the customer and increase the accuracy of the forecast of total supply chain demand as well as the efficiency of supply allocations. Industrial customers usually provide their supplier with ADI and orders through an electronic data interchange (EDI) interface, which is connected to the supplier's ERP system. Historical data is usually stored in a data warehouse backbone of the ERP system. Note that the EDI interface allows customers to update their demand data on a high frequency. In industrial settings, many customers update their data every day. Hence, the analysis of data as well as the planning processes depending on it must be re-run on this high frequency as well.

Another data source for demand planning, supply network planning, and demand fulfilment are the contracts closed between supplier and customers. These define, for example, minimum OLTs, minimum and maximum order quantities, requirements regarding the accuracy of ADI, product prices, obligations regarding stock keeping and minimum service levels, and substitute products. Note that in current industry practice contract clauses are not maintained in machine readable form so that they cannot be used in supply network planning processes in an automated way. Many companies, however, are starting to establish machine readable contract databases containing relevant contract information for demand fulfilment.

In supply network planning, shop floor data is used to plan supply. Usually, the data is maintained by the production sites and stored in databases for master data and databases for dynamic shop floor data. In the master data databases, e.g., resource consumption factors and BOM information of all materials and products is stored. Databases for dynamic data contain information on current cycle times and yields of products on machines or groups of machines, which depend on the current work in progress situation of the shop floor as well as other external, dynamically changing factors.

3 Related literature

Parts of this chapter base on the literature discussions in

Seitz, A., Ehm, H., Akkerman, R., Osman, S., 2016a.
A robust supply chain planning framework for revenue management in the semiconductor industry.
Journal of Revenue and Pricing Management, 15(6), 523-533.

Seitz, A., Grunow, M., 2017.
Increasing accuracy and robustness of order promises.
International Journal of Production Research, 55(3), 656-670.

Seitz, A., Grunow, M., Akkerman, R., 2016b.

Data Driven Supply Allocation to Individual Customers Considering Forecast Bias.

Available at SSRN: https://ssrn.com/abstract=2813835.

The research presented in this thesis is related to the literature streams revenue management, inventory rationing, due-date assignment and scheduling, supply chain coordination with contracts, supply network planning considering ADI, and demand fulfilment in APS. In the following, each of these fields will be discussed.

3.1 Revenue management

The literature on revenue management is divided in quantity-based and price-based approaches (Talluri and van Ryzin 2004). Quantity-based revenue management segments customers by exploiting their heterogeneity regarding, e.g., strategic importance, profitability, price-, or time-sensitivity. Price-based approaches, on the contrary, maximise revenue by actively steering and influencing demand using price differentiation.

Smith et al. (1992), Harris and Pinder (1995) and Ruff (2014) give an introduction to revenue management practices in the airline and manufacturing industries. Quante et al. (2009b) review the literature on revenue management and APS with regards to proposed models, industries and available software. Based on the review, they develop a framework that considers both streams of literature and derive future research opportunities.

An extensive number of studies deals with *quantity-based revenue management*, in which future or current production outputs, inventories, or capacities have to be allocated to different segments, products, or channels (see e.g. Talluri and Van Ryzin 2004, Talluri et al. 2008, Karabuk and Wu 2003, or Huefner and Largay 2013).

In the literature focussing on *quantity-based approaches for industrial environments*, some publications study single-period capacity rationing problems with two customer classes (see e.g. Balakrishnan et al. 1996 or Chiang and Wu 2011). Chen (2006) uses pseudo order information in a single period, single product environment to study the influence of demand uncertainty on order acceptance decisions. The author finds that excess materials increase with rising demand uncertainty while overall profitability of the supply chain declines.

Recent work almost exclusively analyses multi-period cases with multiple customers to derive heuristic or optimal approaches for the calculation of capacity reservations or rules for order acceptance (see e.g. Barut and Sridharan 2004, Hung and Lee 2010, or Chevalier et al. 2015, Pibernik and Yadav 2008, Modarres et al. 2012, Hung et al. 2014). Pibernik and Yadav (2009) nest supply reservations and find that nested reservations have a positive effect on customer service levels. Their numerical study shows further that not knowing the true customer demand has a negative effect on the overall system performance. The approaches presented in Chapters 6 and 7 nest the supply reservations for customers and customer classes and set incentives for the customers to forecast their demands truthfully.

There is also a stream of literature focussing on *quantity-based approaches in the semiconductor industry*. It is mentioned here due to the closeness of this thesis to this industry. The publications mostly present capacity planning and production planning approaches(see e.g. Huang 2005, Zhang et al. 2004, or Chien et al. 2013). The publications confirm that, due to the uncertainty of demands being typical for this industry, planning decisions have to be done in a rolling horizon manner. For this reason, all methods presented in this thesis base on a rolling horizon scheme.

In the field of *price-based revenue management approaches*, Gallego and Van Ryzin (1994), Gallego et al. (2006), Gallego and Stefanescu (2009) and Gallego and Talebian (2012) investigate dynamic pricing problems for price-sensitive and stochastic demand. Using intensity control theory, bounds, and heuristics, they analyse a range of inventory pricing problems. Such approaches can be implemented in the demand fulfilment framework presented in Chapter 4. The publications of Adida and Perakis (2010) and Davizón et al. (2010) put the aforementioned studies in an industry context. They show that the presented approaches enable an earlier revenue stream for manufacturers. Applying the approaches in the semiconductor industry, Constantino et al. 2015 show that a joint dynamic pricing and inventory control model, like the one presented in Chapter 4, can reduce demand uncertainty and the bullwhip effect in supply chains.

3.2 Inventory rationing

Inventory rationing approaches model future demand as a stochastic parameter and allocate inventory to customers with the objective to minimise stock out cost, maximise short-term profits or meet service level requirements. Assuming stochastic, deterministic, or no replenishment lead times, inventory replenishment orders are usually part of the planning decision (see e.g. Ha 1997, Bassok et al. 1999, de Véricourt et al. 2002, Deshpande et al. 2003, Frank et al. 2003, Quante et al. 2009a, Ioannidis 2011, Pinto 2012, Chew et al. 2013, Hung and Hsiao 2013, Wang et al. 2013a/b, or Enders et al. 2014, Han et al. 2014, Pang et al. 2014, Wang and Tang 2014, or Liu et al. 2015). Samii et al. 2011 and Samii et al. 2012 study approaches employing nesting policies and find that nesting leads to higher service levels for high priority demands. Therefore, a nesting strategy is employed in the approaches presented in Chapters 6 and 7.

Recently, some authors have brought inventory rationing literature into the context of supply chain management. Chen et al. (2011) study a system of two revenue maximising physical retailers that serve as drop shippers for an online retailer. The publication analyses the value of information sharing between the retailers and shows that information sharing increases the performance of the supply chain significantly. The approaches presented in this thesis consider ADI shared by the customers, which is typical for industrial settings.

3.3 Due-date assignment and scheduling

Due-date assignment and scheduling approaches quote production completion dates for production orders. Usually a given sequencing policy for the jobs on machines is assumed and the current state of the shop floor, e.g. its congestion, is considered. The publications of Cheng and Gupta (1989), Kaminsky and Hochbaum (2004), Gordon et al. (2012) and Janiak et al. (2015) provide reviews of the field.

The vast majority of publications deals with so-called *static job shop environments*, in which all customer orders are known in advance. Hence, the studied problems are comparable to batch order promising cases from the demand fulfilment in APS literature (see e.g. Dumitrescu et al. 2014, Li et al. 2011, Steiner and Zhang 2011, Yin et al. 2014, Baker and Trietsch 2015).

In the *dynamic job shop problem*, customer orders arrive within the planning horizon (see e.g. Moses et al. 2004, Kaminsky and Lee 2008, Lee 2010, Reindorp and Fu 2011, Slotnick 2011 or Nguyen and Wright 2014). These problems can therefore be compared to the real-time order promising scenarios, this thesis is concerned with.

Only very few publications put due date assignment and scheduling problems into a *supply chain context*. Jin et al. (2013) study a static problem consisting of one manufacturer and several customers where the demand quantity of the customers depends on the assigned due date. The objective is to maximise profit generated by exploiting heterogeneity of customers regarding product price and delivery time. These customer characteristics are similar to aspects described in Chapter 4 of this thesis. Kaminsky and Kaya (2006, 2008, 2009) look at multi-item supply chains consisting of one manufacturer and one or several suppliers in a stochastic environment with dynamic order arrivals where stocking point, stock height, online due date assignment and scheduling decisions have to be made simultaneously. They find that cost can be cut significantly in a combined MTO/MTS system, when supply chains are coordinated centrally or relevant information is shared between the manufacturer and the suppliers. Such environments are dealt with in this thesis (see e.g. Chapter 5).

3.4 Supply chain coordination with contracts

Contracts coordinate the material, value, and information flow in a decentrally planned supply chain. Cachon (2003) provides a broad overview. Tsay et al. (1999) classify contracts into eight categories. For this thesis, which analyses the interaction of the length of OLTs and the accuracy of ADI in demand fulfilment in Chapter 7, the three categories of contracts on quantity flexibility, allocation rules, and lead times are relevant.

In contracts setting *quantity flexibility*, the maximum allowed deviation of the ADI from the final order of a customer and according monetary penalties are determined for different ADI horizons. Contracts setting minimum order *lead times* for customers determine the minimum horizon with which a customer has to place an order to its supplier. The literature in these fields investigates the necessary conditions for supply chain partners to conclude such contracts and determines the optimal behaviour of supply chain partners for a given contractual relationship in order to study its effects on the overall supply chain performance and the distribution of risks within the supply chain (Iyer and Bergen 1997; Tsay 1999; Tsay and Lovejoy 1999; Barnes-Schuster et al. 2006; Lutze and Özer 2008; Kim 2011; Kremer and Wassenhove 2013; Kim et al.

2014; Knoblich et al. 2015). The publications typically assume enough supply to fulfil the total of the demand of the customer.

For supply shortage situations, contracts setting the supply *allocation rules*, i.e. the process of distributing scarce supply between different customers, are needed. Publications in this field typically investigate the effects of such rules on their suitability to coordinate the supply chain efficiently by, e.g., incentivising customers to provide truthful ADI or increasing or decreasing overall profits, costs, stocks, and service levels (Cachon and Lariviere 1999a, 1999b; Plambeck and Taylor 2007; Xiao and Shi 2016; Huang et al. 2013). Many of the studied allocation rules are based on the customer demand forecasts.

3.5 Supply network planning considering advance demand information

When ADI is shared in a supply chain, customers forecast future orders to their immediate upstream supplier a certain time in advance of the order due date. The horizon, over which a customer has to provide forecasts, as well as their maximum variability, are determined in contracts between customers and suppliers. In the investigated system of this thesis, customers forecast their demands using ADI.

The concept of ADI is introduced by Hariharan and Zipkin (1995) and Buzacott and Shanthikumar (1994) as an effective means to reduce uncertainty in industrial supply chains with exogenous and endogenous supply lead times. Most publications in this field study the influence of unbiased or biased ADI on the customer order decoupling point for given production control strategies (Karaesmen et al. 2002; Claudio and Krishnamurthy 2009; Karrer et al. 2012; Altendorfer and Minner 2014). Ouyang and Daganzo (2006) and Ouyang (2014) show that ADI can reduce, but not eliminate, demand volatility and uncertainty in supply chains. Therefore, it is necessary to develop operational means that exploit supply chain flexibilities to cope with the remaining demand uncertainty, take the heterogeneous lead times resulting from the ADI into account, and incentivise customers to provide unbiased ADI. Such approaches are presented in this thesis.

3.6 Demand fulfilment in advanced planning systems

The publications dealing with demand fulfilment in APS can be categorized in literature describing applications for homogeneous and heterogeneous customer OLTs. The former category can be detailed further into contributions for order promising, order re-promising, and their combination, so-called hybrid approaches. After giving an overview of literature dealing with general requirements for and applications of demand fulfilment methods (Section 3.6.1), the literature from each category is reviewed in Sections 3.6.2 and 3.6.3.

3.6.1 Categorisation and conceptualisation of approaches

Fleischmann and Geier (2012) and Kilger and Meyr (2015) give a general introduction to the tasks and concepts of demand fulfilment. They present models from the literature dealing with order promising and allocation planning. Further, the influence of the location of the customer order decoupling point on the demand fulfilment process is discussed.

Ball et al. (2004) name several other dimensions and factors that affect the demand fulfilment process. The authors state that the robustness and accuracy of order promises, which is a focus of this thesis, is of high importance, especially in industrial environments.

Pibernik (2005) proposes a theoretical framework for demand fulfilment, called advanced available-to-promise. The author identifies eight generic advanced available-to-promise types and three additional functionalities of advanced available-to-promise approaches. Accordingly, the demand fulfilment methodologies presented in this thesis can be categorised as passive real-time multi-location approaches for finished good supply allowing partial deliveries.

Framinan and Leisten (2010) identify order acceptance and selection, due date assignment and order scheduling as the main decisions of demand fulfilment. They derive eight types of demand fulfilment approaches, of which the methods presented in this thesis can be seen as Approach IV: integrated order acceptance and due date assignment.

3.6.2 Methods for homogeneous customer order lead time environments

Models for homogeneous customer OLT environments are divided into approaches for order promising, order re-promising, and hybrid methods. To structure the vast amount of literature presenting order promising approaches, this category is further divided into methods for MTS, assemble-to-order (ATO), and MTO supply chains.

Order promising models are further divided into batch and real-time approaches. Batch order promising approaches imply customer prioritization possibilities and are therefore usually operated without preceding allocation planning. They assume that customers are willing to wait until they receive an order promise, because orders need to be collected over a certain amount of time, i.e. the batching interval. In industrial environments, however, short customer response time is perceived as good customer service. In such cases, real-time order promising approaches need to be employed, which require preceding allocation planning mechanisms. Therefore, the methods presented in this thesis employ allocation planning and real-time order promising processes.

3.6.2.1 Allocation planning

In practice, common *allocation planning* approaches still use simple business rules (Kilger and Meyr 2015, Cederborg and Rudberg 2009, Pibernik 2006) even though they are known to increase the bullwhip effect (see e.g. Bakal et al. 2011). However, many more sophisticated approaches can be found in the literature. Alarcón et al. (2009), Lečić-Cvetković et al. 2010, Babarogić et al. (2012) and Ali et al. (2014) propose procedural frameworks that include supply allocation and aim at maximising customer service levels or short-term profits. Other authors study stylised special cases using probabilistic modelling to derive algorithms (Pibernik and Yadav 2009) or structural characteristics of the optimal order acceptance policy (Chiang and Wu 2011, Gao et al. 2012, Papier 2016). Many scholars present linear programming-based approaches in which the aim is to maximise overall profits by integrating AP with production planning in assemble-to-order or make-to-order environments (Ball et al. 2004, Ervolina et al. 2009, Chen and Dong 2014, Chiang and Hsu 2014) or in make-to-stock environments (Huaili and Yanrong 2010, Meyr 2009, Lebreton 2015, Alemany et al. 2015). Vogel (2014) proposes a method for multi-stage customer hierarchies. He shows that the approach can lead to higher profits compared with profits achieved by an optimal central allocation approach, if demand forecast accuracy is very low.

Meyr (2009) presents linear programming models for allocation planning in MTS environments. In a case study from the lighting industry, the approach is compared with conventional approaches with and without customer segmentation. The results show that customer segmentation and allocation planning leads to a substantial increase of profits, if customers are heterogeneous and the information on available supply and customer demand is accurate. Another numerical finding is that the number of customer classes influences the performance of the approach more than the segmentation method. Parts of the presented demand fulfilment processes in Chapters 5 to 7 build on this work, extending and modifying the models presented in Meyr (2009).

3.6.2.2 Order promising in make-to-stock supply chains

In industrial environments, mostly rule-based MTS approaches searching through different dimensions of ATP like product, time, customer group, or region are used to promise orders in real-time (see e.g. Kilger and Meyr 2015, Pibernik 2005, Fleischmann and Geier 2012, or Ball et al. 2004).

Pibernik (2005) presents a linear programming based batch order promising method. The model maximises profit considering revenue created through order promises, inventory holding, handling, and shipping costs, and penalty costs for order rejection. The publication mentions possible extensions for latest promising dates, multiple partial deliveries, and multiple locations. However, no mathematical representations of these extensions are provided.

Jung (2010) proposes a linear programming model for batch order promising. The approach integrates predefined customer priorities as well as earliness and tardiness cost of order fulfilment.

Meyr (2009) presents linear programming models for allocation planning and real-time order promising, which maximise company profits. In a numerical study, performance of the approach is compared to a batch order promising model without ATP allocations. The works of Yang (2014) and Eppler (2015) build on this publication and develop probabilistic approaches that are scalable to industrial problem sizes and outperform the approach presented in Meyr (2009) in terms of revenue.

3.6.2.3 Order promising in assemble-to-order supply chains

In ATO environments, the ATP information consists of supply for components that are assembled to finished products upon customer order arrival. Dickersbach (2009) describes a rule-based real-time order promising approach for material-constrained ATO environments, in which the availability of components is the main bottleneck for the capability of the supply chain to produce requested goods. When an order arrives, the availability of the needed components is checked separately. Afterwards, a fixed production lead time is added to the latest component availability date. Dickersbach reports that the approach leads to inaccurate delivery dates, if it is applied to complex real-world production environments.

Cederborg and Rudberg (2009) describe the demand fulfilment process of a steel manufacturing company with divergent material flows. The rule based order promising approach is similar to that described by Dickersbach (2009). However, if upon order arrival certain components are not available, the standard production lead time for these components is used to promise the order. The authors mention that the order promising process has to take dynamic routing decisions into account. However, the meaning of this term remains unclear. Additionally, no quantitative results are presented in the publication.

Tsai and Wang (2009) present a three stage mixed integer linear programming model for batch order promising. In the first step, orders are assigned to assembly plants on basis of coarse availability information per plant. In the second step, fine grain planning determines the order promises for every plant individually. Orders that cannot be satisfied in the second step are assigned to other plants in the third step.

Lin et al. (2010) suggest a batch order promising model for the TFT-LCD manufacturing industry that takes customer individual profits as well as capacity and material constraints of production sites into account. Gössinger and Kalkowski (2015) present a similar approach that aims at providing profitable and reliable delivery date promises by making use of three preventive measures. The approach is validated using a case from the customized leisure products industry.

3.6.2.4 Order promising in make-to-order supply chains

In MTO environments, production is only started after the customers submitted their orders. The ATP information therefore usually consists of data on material and capacity availability. Jeong et al. (2002) propose a greedy algorithm for batch order promising in the electronics industry. The approach promises orders based on finished product supply in distribution centres as well as idle capacities in the shop floor, which are derived from the current master production schedule. In the batch, orders are prioritized based on their arrival time, order quantity, tightness of due date and customer priority.

Dickersbach (2009) describes two real-time order promising methods for solely capacity-constrained cases. The approaches use the supply network planning process to promise orders. The order delivery dates are derived by inserting the newly arrived order into the existing production schedule. Thereby, promised dates of already promised orders must not be violated but their capacity consumptions can be rearranged. In the literature, such approaches, which use the supply network planning process to promise orders, are also called capable-to-promise (CTP). Dickersbach states that such approaches can cause scattered capacity loading. Therefore, they require frequent demand supply matching processes that optimise capacity utilisation. Such frequent re-planning activities are also employed in the methods presented in Chapters 4 and 5.

Moses et al. (2004) develop a CTP approach that reflects the variance of production lead times for individual orders. In a numerical study, the robustness of the calculated order promises is measured in terms of order tardiness and absolute lateness. Rabbani et al. (2014) develop a genetic algorithm for integrated order promising and scheduling in a multi-machine flow shop production environment. Jung (2012) suggests a fuzzy linear programming solution calculating delivery dates which are sent to the customers for negotiation. Brahimi et al. (2014) propose two heuristics for a mixed integer linear programming approach for integrated production planning and order acceptance decisions for order batches. Yang and Fung (2014) present two order promising solutions integrating order acceptance, due date assignment and order scheduling in a multi-site supply chain.

3.6.2.5 Order re-promising

Fleischmann and Meyr (2004) present basic linear programming models for order re-promising in MTS, ATO, and MTO environments. They also describe several extensions for shortage planning.

For a case of a mobile phone manufacturer, Klein (2009) proposes a mixed integer linear programming model for an ATO environment. The model decides which orders are not produced in case of a shortage situation. The approach penalises due date violations. Interdependencies with the order promising process are not investigated.

3.6.2.6 Hybrid order promising approaches

Ball et al. (2004) describe a hybrid order promising process for the ATO case on the example of the computer manufacturer Dell. Upon customer order arrival, a coarse online promise with a preliminary delivery date is given. These initial order promises are refined in a batch re-promising process about 14 days after order reception. The refined order promise can deviate from the initial promise. No mathematical model of the approach is presented.

Geier (2014) develops another hybrid order promising approach for another ATO case from the computer industry. The presented solution takes substitution of components and alternate sourcing as shortage planning approaches into account. It is shown that problems of realistic size can be solved with mixed integer linear programming approaches. The publication is the only one investigating the interactions between real-time order promising and re-promising.

3.6.3 Approaches for heterogeneous customer order lead time environments

Building on previous research published in Chen et al. (2001), Chen et al. (2002), and Ball et al. 2003, Zhao et al. (2005) present an optimisation model for a batch order promising case from the electronic product manufacturer Toshiba. The order promising horizon is divided into three partitions, in each of which different ATP information is used to calculate order promises. In the near-term, incoming orders are promised on finished product ATP quantities. In the mid-term, order promises are calculated on the basis of component and capacity ATP. In the long-term, only capacity ATP is used. The publications are the only contributions describing a partitioned order promising process, i.e. order promising in an environment with heterogeneous customer OLTs.

Kaminsky and Kaya (2008, 2009) investigate multi-item supply chains in a stochastic environment with dynamic order arrivals. They find that cost can be cut significantly when MTO and MTS systems are combined and relevant information is shared between supply chain partners.

3.7 Delimitation of this thesis from the existing literature

The above literature review shows that there is a vast number of contributions, which applies concepts from manifold disciplines in the individual processes described in Section 1.2. However, so far a supply chain planning framework that combines the different aspects addressed, balances demand and supply, and increases demand fulfilment robustness and accuracy using the different ideas from all mentioned literature streams, is missing. In Chapter 4, this gap is filled through the proposal of such a framework. It provides a coherent structure for the above-mentioned literature, as well as an instrument that facilitates the implementation of current research and the identification of future research directions. By aligning the planning objectives of the different supply chain planning processes and exploiting newly available data, the framework improves demand fulfilment performance of industrial supply chains.

Revenue management (see Section 3.1) and inventory rationing (see Section 3.2) methodologies typically include supply replenishment decisions into their allocation of availabilities. As described in Section 1.2, this thesis investigates common industrial environments, in which the planning hierarchy separates supply decisions from allocation decisions. In this context, allocation decisions have to be made based on exogenously given ATP supply. Moreover, revenue management and inventory rationing approaches differ from the methods presented in this thesis in terms of the type of resources, which are considered in the planning decision. While revenue management approaches allocate perishable availabilities, i.e. production capacities, to different customer classes, inventory rationing methodologies usually distribute durable resources, i.e. finished products. This thesis, in contrast, considers both durable and perishable resources (see Chapter 5).

Another, differentiating factor is that revenue management and inventory rationing methodologies usually assume stochastic demand. This requires the knowledge of the probability distribution functions of future demand, which is usually not the case in industrial environments because the planning tools used do not support the derivation of probability functions but use internal or external demand forecasts in demand fulfilment typically coming from sales and marketing functions or customers. Quante et al. (2009a) find that their stochastic approach outperforms deterministic optimisation methods, like the ones presented in this thesis, for high demand forecast errors. However, the presented solution is not scalable and cannot be applied to larger problem sizes often found in industrial practice.

Due date assignment and scheduling (see Section 3.3) approaches mainly focus on scheduling a given or uncertain amount of production orders on production resources on the shop floor level. This thesis, however, deals with problem settings on a supply chain-wide level. Further, customer prioritization and reservation of capacities for future order arrivals are typically not discussed in this field.

Most of the reviewed publications on customer contracting aim at coordinating the supply chain globally by using game theoretical approaches. Apart from the work on allocation rules, the previous research focuses on dyadic contractual relationships between a supplier and a customer, for which it is decided which contract terms to include and how to set them. The same contract terms apply to all customers. An exception is Barnes-Schuster et al. (2006), who determine order lead times for multiple customers minimising global supply chain cost. However, it is assumed that the supplier can always fulfil the total demand of customers within the contractual order lead time. Furthermore, the exchange of ADI is not considered in the publication. In general, the effects of allocation rules, ADI terms, and order lead time contracts have been addressed, but only separately, neglecting their interactions in the structure and dynamics of the demand fulfilment process. Moreover, the stylized analyses do not reflect the structure and dynamics of the demand fulfilment process. This, however, is necessary to investigate the interaction of ADI and OLTs on the demand fulfilment performance in supply shortage situations. The research presented in this thesis, in contrast, takes the perspective of the supplier. It aims at drawing conclusions on how a supplier should design the entire portfolio of contracts of all customers, in which the terms may be different for different customers. The implications of different contractual designs are investigated in the dynamic context of a new demand fulfilment approach, designed to be incorporated in the structure of industrial planning environments (see Chapter 7).

This thesis differs from the reviewed literature on supply network planning considering ADI (see Section 3.5) because it investigates the influence of the design of contract portfolios determining order lead times and ADI on the supply chain performance for a given customer order decoupling point and demand fulfilment process. The relevant supply chain planning decision considering ADI is not the supply network planning strategy, i.e. the decision between

MTS, ATO, or MTO, but the distribution of scarce supply to customers considering the relationship of OLT and ADI in the set of customers (see Chapter 7).

The majority of existing publications on demand fulfilment in APS (see Section 3.6) presents models for homogeneous customer OLTs. Only the publications of Zhao et al. (2005) and Kaminsky and Kaya (2008, 2009) study environments with heterogeneous customer OLTs. The approach presented in Zhao et al. (2005) shows similarities to the one presented in Chapter 5, since it considers heterogeneous customer OLTs. Moreover, supply chain flexibilities can be exploited in the mid- and long-term horizons, but for short-term customer orders order promises are calculated based on a fixed finished product production plan. Also, the ATP type used for the mid-term planning horizon shows similarities to the cumulated ATP (CATP) presented in Chapter 5. However, Zhao et al. (2005) only mention the usage, but do not show how to generate their mid-term ATP. Also, it remains unclear how this ATP type is consumed when orders are promised. Therefore, this thesis is so far the only one showing how to calculate ATP information that can be used to exploit supply chain flexibilities. Furthermore, the lengths of the partitions of the order promising horizon are fixed in Zhao et al. (2005). The approach presented in Chapter 5 uses product individual production cycle times in order to reflect the reality of different cycle times of different products. Additionally, instead of using three different types of ATP and order promising models in a partitioned planning horizon, the approach presented in Chapter 5 only uses one type of ATP and one promising model. The approach is therefore of higher practical usability.

Solely the work of Geier (2014) considers the robustness of given order promises and penalises updates of order promises in the order re-promising step. Means to increase the robustness of order promises using the real-time order promising process are not addressed. Therefore, the approach does not increase the accuracy of order promises. Chapter 5 fills this research gap by presenting an order promising approach that increases the accuracy and robustness of order promises by making use of ATP information that considers supply chain flexibilities.

CTP approaches for demand fulfilment are reported to perform poorly in complex industrial environments (see e.g. Dickersbach 2009 or Quante et al. 2009b). Their order promises are either unreliable because the aggregation level used for the planning parameters is too high, or their computation times are prohibitively long when planning is done on a more detailed level. Batch order promising approaches lead to scattered production schedules that necessitate regular order re-promising (see Dickersbach 2009). They are computationally expensive and of myopic nature since they do not take future orders into account (see Pibernik 2005). They also assume that customers are willing to wait until orders promised, because orders need to be collected into batches over a certain amount of time, i.e. the batching interval. In industrial environments, this is often not the case since short customer response time is perceived as good customer service. In such cases, real-time order promising approaches are needed, which are investigated in this thesis.

Framinan and Leisten (2010) mention five flexibilities that can be exploited when it is not possible to fulfil orders according to the original request. With product (substitution) flexibility, the supplier can ship substitute products as an alternative to the originally ordered product. Volume flexibility provides the possibility to fulfil orders by partial shipments. Delivery flexibility enables postponement of the order due date. Resource flexibility allows the use of alternative sources for one demand, e.g. finished goods inventories or manufacturing capacities. When price flexibility exists, the supplier is able to charge a higher price to make order acceptance profitable. This thesis, in contrast, identifies two new supply chain flexibilities that can be exploited in demand fulfilment. Here, product flexibility is defined as the possibility to produce different products from one intermediate product or raw material and process flexibility is the possibility to use one process for the production of different products.

Although Chiang and Hsu (2014) mention the use of order behaviour information in demand fulfilment, none of the mentioned demand fulfilment approaches consider the accuracy of ADI provided by customers or the customer OLT in their planning models. Furthermore, all reviewed demand fulfilment publications allocate supply to customer segments in a mid-term rolling horizon fashion. In contrast to that, the allocation planning approaches presented in Chapter 6 and 7 explicitly use ADI bias and OLT data as two examples of order behaviour information and allocate supply to individual customers in a short-term rolling horizon manner.

Finally, Papier (2016) shows that the usage of ADI in allocation planning to different markets can increase the expected profit significantly. In contrast to the methods presented in this thesis, re-allocation of supply is not possible in the approach. Furthermore, as in this thesis supply is allocated to individual customers, the approaches are designed for a different level of aggregation.

4 A data driven framework for robust and accurate demand fulfilment

This chapter bases on

Seitz, A., Ehm, H., Akkerman, R., Osman, S., 2016a.
A robust supply chain planning framework for revenue management in the semiconductor industry.
Journal of Revenue and Pricing Management, 15(6), 523-533.

In this chapter, a data driven framework for robust and accurate demand fulfilment in industrial environments is presented. It consists of robust and flexible solutions for demand steering and dynamic pricing, extending current industry practice in several aspects. The concept of availabilities and capabilities (A&C), as well as various planning processes and process enablers are introduced. The framework is developed on the example of the semiconductor industry, in which high demand uncertainties, long production lead times and short product life cycles cause high risks for supply chain planning. Such characteristics can be found in various other industrial environments. Hence, even though it is developed for the semiconductor industry, the framework can be applied in many other manufacturing industries as well. Based on the framework, directions for future research are highlighted.

In the following, Section 4.1 gives an overview of the challenges for supply chain planning in the semiconductor industry. Sections 4.2 to 4.6 describe the framework in detail and contrast it with current best practices in the semiconductor industry. Sections 4.7 and 4.8 discuss future research directions based on the framework and present conclusions, respectively.

4.1 Consequences of operational inflexibility in volatile markets

To remain competitive in the global markets of today, companies continue to specialize on their core competencies. Consequently, the number of partners interacting in global supply chains and the complexity of company networks rise constantly. Therefore, supply chains become increasingly sensitive to disturbances caused by the manifold interactions between their entities. Human- as well as system-caused misalignments lead to distortions in forecasts and order management processes. In the semiconductor industry one of these distortions, the so-called bullwhip effect (see e.g. Lee et al. 2004), causes severe amplifications of demand signals in the supply chain. For example, the global semiconductor market without memory and microprocessors shrank by almost 40% in 2009 while it grew over 40% in 2010 (WSTS Inc. 2015). Similar distortions are to be expected in the future as well. Since semiconductors are present in almost all products used by industrial and end customers, the bullwhip effect therefore poses a significant problem of modern society. Other challenges for the industry are very short customer OLTs (the time interval between order entry and requested delivery date) and low accuracy of the ADI, i.e. demand forecasts, provided by customers.

However, the operational flexibilities of semiconductor manufacturers to react to such uncertainties are limited. On the one hand, long production cycle times of up to five months force manufacturers to start production long before their customers place their orders. On the other hand, due to the well-known Moore’s law (see e.g. Moore 1965 or Bergeron 2008), innovation is fast. In consequence, life cycles of semiconductor products are short and, hence, the risk of obsolescence of stocks is high. Thus the possibilities for semiconductor manufactu-

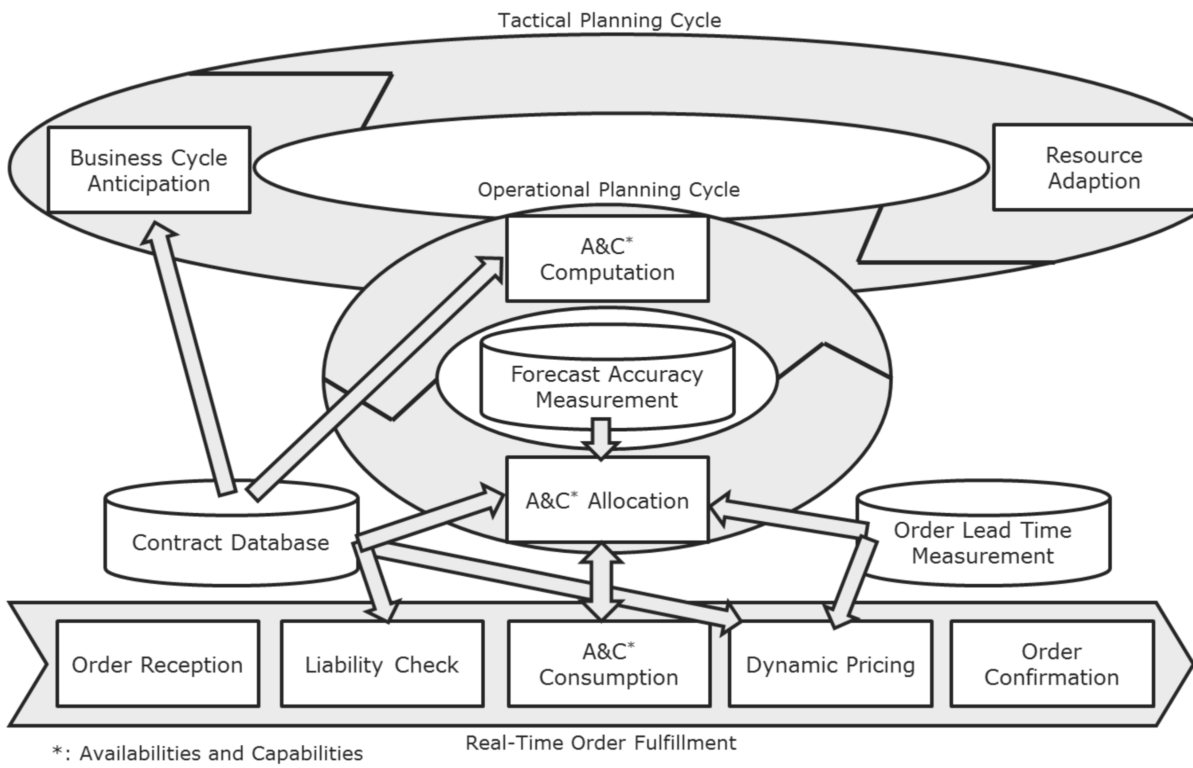

Figure 3: Data driven supply chain planning framework

rers to adapt production volumes or buffer stocks to react to short term demand changes are limited. Furthermore, since the economic cycles are becoming shorter, periods of supply shortage and high excess supply are not only more severe but also appear more frequently. To keep utilizing capacity in a profitable way, semiconductor companies have started to exploit new ways to balance demand and supply by using ideas from various scientific literature streams such as revenue management, inventory rationing, and demand fulfilment in APS.

In this chapter, a framework for robust and accurate supply chain planning in the semiconductor industry is presented. The basic idea of the approach, employing robust and flexible solutions, is twofold. First, it aims to even out demand maxima and minima, steer demand to otherwise unused capacities, and fulfil unforeseen short term demands by means of business cycle anticipation, supply chain resource adaptation, supply allocation, as well as dynamic pricing processes. Second, it aims to maximise revenue and build financial capabilities to maintain higher operational buffers leading to higher customer service through differentiated pricing for urgent orders. The framework outlines the current best practice in supply chain planning for the semiconductor industry while also being the first to apply robust revenue management in manufacturing practice.

4.2 Supply Chain Planning Framework For Dynamic Pricing And Demand Steering

Figure 3 shows the framework. It consists of:

- a tactical planning cycle (business cycle anticipation and resource adaptation),
- an operational planning cycle (A&C computation and A&C allocation),
- five real-time demand fulfilment processes (order reception, liability check, A&C consumption, dynamic pricing, and order confirmation), as well as

- three process enablers (a machine readable contract database, automated forecast accuracy measurement, and OLT measurement).

Following traditional order management concepts such as ATP and CTP (see e.g. Pibernik 2005), availabilities are defined as current inventories and future production that is planned based on demand information consisting of current orders on hand as well as company internal and customer demand forecasts. Similarly, capabilities are defined as possible production quantities that can be realised with idle resources in the current master production plan if demand can be generated. Note that capabilities are not produced if order management processes do not generate demand for them.

In the following, each of the elements of the framework, their interactions, and their main contributions to best practice in the semiconductor industry are described in detail.

4.3 Process Enablers

The framework contains three process enablers that are a novelty in industry practice. The contract database contains demand fulfilment-relevant information from supply agreements between the company and its customers in machine readable form. In current industry practice companies usually define the terms and conditions of doing business with each of their customers by means of such contracts. However, the information contained in these contracts is not maintained in a way that it can be used for automated demand fulfilment processes. In the framework, information from this database is used in the processes business cycle anticipation, A&C computation, A&C allocation, liability check, and dynamic pricing.

The purpose of the forecast accuracy measurement is to measure and statistically analyse the historical accuracy acc_{ih}^{fc} of demand forecasts, i.e. ADI, provided by customer i depending on the horizon h between forecast entry and forecasted demand due date. For the calculation of acc_{ih}^{fc}, Equation (37) can be used, in which err_{ih}^{fc} is defined as the historical forecast error of customer i for horizon h. For the calculation of err_{ih}^{fc}, SMAPE (see e.g. Armstrong 1985 or Ott et al. 2013) is used, which is defined in Equation (38). Here, $f_{it\tau}^{hist}$ is the forecasted order quantity provided by customer i in period t with requested delivery date in period τ, $q_{i\tau}$ is the realised order quantity received from customer i with requested delivery date in period τ, and T^{hist} is the number of observed time periods.

$$acc_{ih}^{fc} = \left(1 - err_{ih}^{fc}\right) \tag{37}$$

$$err_{ih}^{fc} = \frac{\sum_{\tau=1}^{T^{hist}} \left| f_{i(\tau-h)\tau}^{hist} - q_{i\tau} \right|}{\sum_{\tau=1}^{T^{hist}} \left(f_{i(\tau-h)\tau}^{hist} + q_{i\tau} \right)} \tag{38}$$

The OLT measurement provides statistics on the historical customer OLT, i.e. the historical difference between order entry and requested delivery dates. To identify the distribution of OLTs, the Kolmogorov-Smirnov test (Massey 1951) or the Anderson-Darling test (Anderson and Darling 1954) are used.

In other words, the forecast accuracy measurement and OLT measurement processes quantify the uncertainty of customer demands. The measurements are done on final product level. Data from both enablers is used in the A&C allocation process, whereas only the data from OLT measurement is used in the dynamic pricing process.

4.4 Tactical planning cycle

The purpose of the two tactical planning processes (business cycle anticipation and resource adaptation) is to forecast future customer demand for the mid- and long-term horizon, and to adapt production resources accordingly. These semi-automated processes require decision-making on higher management levels and provide necessary inputs for the subsequent processes in the framework. The forecasting of future demand is based on external and internal sources. External sources are e.g. predictions of the economic cycle of public or private research institutes, press releases of competitors and companies in related industries, and political and economic coverage in media. Internally, classical sources like reporting from operations, sales, and marketing, but also novel approaches like crowd opinion or cloud data analysis are used. Production resources are adapted on basis of the demand forecasts resulting from business cycle adaptation. The biggest challenges in these processes are to align all planning dimensions (i.e. demand, revenue, capacity, and production volumes) and to shorten the planning processes to enable shorter planning cycles. These shorter cycles are needed to cope with the increasing volatility of market environments and enable faster adaptation of resources. Furthermore, it is important to enable decision-making on aggregated planning levels while disaggregation of resulting decisions has to be automated. Here, the key is to identify aggregation levels that enable precise decision-making, while at the same time allowing a global view on the decision problem. Performance indicators used to measure the quality of tactical planning processes are, amongst others, the resulting forecast errors in terms of demands and revenues, as well as resource utilization and work-in-progress in manufacturing sites.

Note that business cycle anticipation and resource adaptation are processes aiming at stability, i.e. constant utilization of supply chain resources during the entire economic cycle. However, due to the aforementioned short-term and mid-term demand uncertainties, long production lead times, long resource ramp-up times, and high investment cost for production equipment in the semiconductor industry, it is usually not possible to fully adapt production resources to demand. Consequently, periods of resource shortage and excess appear even though the above described processes are implemented and run effectively. These risks need to be mitigated with robust and flexible planning processes on the operational level, which are described below.

Additionally, although business cycle anticipation and resource adaptation are current practice, the systematic usage of cloud data, crowd opinion techniques, and contractual data in business cycle anticipation is novel in the industry.

4.5 Operational planning cycle

A&C computation generates the A&Cs (whose sum is called supply in the remainder) that serve as input for the processes A&C allocation and A&C consumption. Standard supply chain planning tools and functionalities are used. First, short- and mid-term demand data is booked into supply chain resources to generate availabilities. Afterwards, to generate capabilities for periods, in which idle capacities occur, the demand information is inflated and the delta is booked into remaining idle capacities until capacity utilization reaches a target limit. In time periods where the capacity is too small to serve all forecasted demand, information from the contract database is used to prioritize demand that the company is obliged to fulfil. A utilization target rather than revenue target is employed for two reasons. First, production resources in the semiconductor industry are highly capital intensive and often a constraining factor for demand fulfilment. Therefore, the utilization of capacities is a key driver of profitability in this industry. Second, due to the dynamic pricing process described below, the revenue that will be generated from the creation of capabilities is hard to predict.

The resulting supply is assigned to the most profitable and strategically important customers in the A&C allocation process. Supply is allocated respecting contractual obligations to the customer, urgent supply needs due to line down threats at customer sites, strategic minimum service levels for customer groups, firm and already committed customer orders on hand, as well as individual customers and total supply available. Furthermore, the allocation fulfils internal demand forecasts and ADI from the customer using target service levels for customer groups and individual customers. The ADI is discounted by the historical bias of the customer demand, which is calculated in the forecast accuracy measurement process. In other words, customer forecast biases stemming from gaming behaviour or systematic differences in the planning systems of customers and the company are eliminated. Customer forecasts are discarded in case their remaining forecast lead time, calculated by the difference of the forecasted delivery date and the current date, is less than the 5% percentile of the OLT distribution of the respective customer, which is calculated in the OLT measurement process.

Additionally, product substitution to increase service levels is considered in the A&C allocation process. Supply exceeding demand forecasts on the individual customer or customer group level is made available to all customers in a customer group or all customers, respectively. By regularly rerunning the A&C allocation process, the allocated quantities are continuously revised to minimise the risk to reserve supply for possible customer demand that is not likely to be realised while incoming orders of other customer groups have to be declined because their allocated quantities are used up.

The operational planning cycle aims to ensure plan stability by calculating the maximum output (in terms of A&C) of the supply chain and constantly allocating this supply to customer groups and individual customers under consideration of contractual obligations and demand uncertainties. Thereby, robustness is achieved by minimising the risk of having to change confirmations of already accepted customer orders. Simultaneously, the generation of capabilities enables flexible demand steering processes in real-time demand fulfilment (see next section). This, as well as the usage of contractual information, forecast accuracy, and OLT information in A&C computation and allocation, extends current best practices in the industry.

4.6 Real-time demand fulfilment processes

The real-time demand fulfilment processes order reception, liability check, A&C consumption, dynamic pricing, and order confirmation are performed every time an order arrives. First, the new order is checked for contract liability. For this, agreements on minimum OLTs as well as minimum and maximum order quantities are retrieved from the contract database. Afterwards, the order is sent to the A&C consumption process which calculates the earliest possible delivery date, disregarding the result of the liability check, using standard order promising functionalities (see e.g. Kilger and Meyr 2015). Thereby, the allocations established by the A&C allocation process are respected.

Note that it is always possible to promise an order with its contractually binding minimum OLT and maximum order quantity since contract information is used in the resource adaptation, A&C computation, and A&C allocation processes.

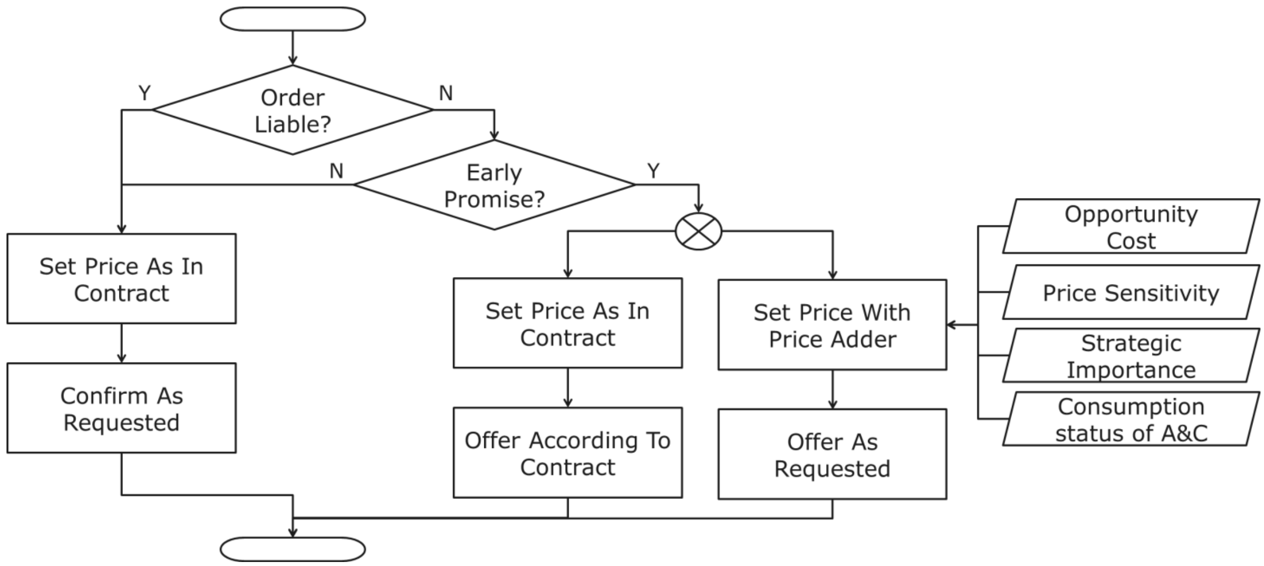

Figure 4: Dynamic pricing approach and order confirmation processes

The subsequent dynamic pricing and order confirmation processes are shown in Figure 4. If the order is liable, the order is confirmed at its requested date with the product price contractually agreed on. Otherwise, if a confirmation earlier than the contractually binding delivery date is possible, two offers are made to the customer. One contains the earliest possible delivery date calculated by the A&C process. The other contains the contractually binding delivery date and quantity retrieved from the contract database. For the first offer, a price higher than the one contractually agreed on is charged. This price depends on the price-sensitivity and the strategic importance of the customer as well as the consumption status of supply and the opportunity cost for consuming supply that could otherwise be used for potentially later incoming orders of more profitable customers. The price of the second offer is the price contractually agreed on. The price of the first offer is supposed to either increase the revenue in the period of the delivery date calculated by A&C consumption or, in case supply in this time period is scarce, steer the demand to the contractually binding time period. In case it is not possible to confirm the order earlier than at the contractually binding delivery date, the order is confirmed with its agreed minimum OLT and price.

The real-time demand fulfilment processes are flexible processes since they steer demand to time periods of low resource utilization, if it does not have to be fulfilled at its requested date because of contractual obligations. Thereby, contractual agreements as well as already accepted customer confirmations are taken into account so that robustness of the approach is ensured. Additionally, the flexible dynamic pricing process creates additive revenues enabling the company to hold higher buffers in the supply chain and hence provide better service to the customers.

Note that in current industry practice, orders are normally promised to the requested or otherwise earliest delivery date possible. Contractual liability of the incoming order is usually not taken into consideration. Therefore, dynamic pricing of orders lying outside of contractual agreements is not possible, which is another aspect in the framework that surpasses current industry practice.

4.7 Future Research Directions

Based on the framework, many directions for future research can be identified. First of all, the existence of the bullwhip effect as well as the mentioned demand uncertainties in current practice show that new and better forecast and optimisation methods are necessary in the business cycle anticipation process of the framework. The mentioned sources of information for this process are a first step towards their development.

Moreover, today, human planners and automated planning processes are mainly working towards local optima for the part of the supply chain or customer segment they are responsible for. This leads to sub-optimal results regarding the performance of the global supply chain. However, changing the focus of planning activities from local to global optimisation is a challenging task. For example, in the A&C computation process, human planners and planning processes should work towards global service level and revenue optimisation considering all customers rather than optimising the utilization of supply chain resources as a main goal. Similarly, all processes of the framework should be oriented towards the ultimate goal to fulfil customer requirements as well as possible while increasing revenues. Establishing the necessary transparency, suitable performance indicators, and tools are interesting fields of research. Simulation techniques combining discrete-event and agent-based modelling, which have recently found their way into semiconductor supply chain management, could help to achieve this goal since they are capable of modelling system as well as human behaviour. Furthermore, a high effort to train human planners must be made to achieve the required robustness in supply chain planning.

Additionally, the tactical processes business cycle anticipation and resource adaptation are currently predominantly done manually. Automated decision support systems for these processes would increase their efficiency and transparency substantially. However, in these decision processes, a multitude of quantitative and qualitative measures, including the gut feeling of experienced planners and managers, have to be taken into account. Tools of cloud data analysis, crowd opinion, and big data analysis are promising technologies to drive automation in these fields.

Also, no supply allocation mechanism fulfilling all mentioned requirements has so far been developed. The main challenge here is to develop an approach that considers demand uncertainty and can be integrated into industry-typical APS, which usually consider demand forecasts to be deterministic. For this, forecast accuracy and OLT measurement need to be established and made compatible with industry-typical IT tools. Such approaches are developed in Chapters 6 and 7.

A rather managerial challenge is to establish machine-readable contract databases and integrate them into current ERP systems. The high efforts of establishing and maintaining such a database makes organizations reluctant to implement such a solution.

Finally, a method implementing the shown dynamic pricing and order confirmation process has to be developed. The challenge is to find industry-suitable models for customer price sensitivity as well as opportunity cost representations for supply consumption.

4.8 Conclusion

In the semiconductor industry, human- as well as system-caused misalignments across the supply chain lead to severe demand uncertainties that have to be dealt with in the supply chain planning processes of a company. In this chapter, revenue management ideas from the service industries are transferred to the semiconductor industry by proposing a data driven supply chain planning framework supporting improved demand management. The framework was developed for a large European semiconductor manufacturer who currently establishes methodologies, algorithms, and processes to implement the framework in its own planning landscape.

The framework aims at robustness by minimising the risk of supply shortage and idle resources by improving the quality of short-, mid- and long-term demand forecasts and adapting and allocating supply chain resources accordingly. This is mainly reflected in the processes of business cycle anticipation, resource adaptation, A&C computation, A&C allocation, OLT, and forecast accuracy measurement.

The framework aims at flexibility by steering incoming demand with uncertain OLTs and order quantities to the most profitable point in time while at the same time mitigating the bullwhip effect. These aspects are mainly covered in the processes of order reception, liability check, A&C consumption, dynamic pricing, and order confirmation.

Many of the elements presented in the framework surpass current best practices in the industry. Firstly, the concept of computing capabilities, i.e. possible production quantities that can be realised with idle resources in the current master production plan, if demand can be generated, and considering them in a continuous allocation process is not part of current practice. Furthermore, the idea of continuous allocation of supply for demand steering and smoothing purposes considering contractual obligations, forecast accuracy, and OLT statistics as well as the processes liability check and dynamic pricing are novelties in the semiconductor industry.

The bullwhip effect is one of the main challenges in supply chain management, and has been extensively studied for many years. Despite this, it has not been significantly reduced yet. Changing human behaviour to overcome it is difficult since it requires collaboration over the whole supply chain. The framework presented in this chapter is a first promising step towards mitigation of demand fluctuations in semiconductor supply chains.

Revenue management has been applied in the service industry for decades. However, the semiconductor and other business-to-business industries did not manage to transfer these ideas into their supply chain planning in the past. Now the introduction of revenue management can be successful due to three reasons: First, new data analysis, automation, and simulation techniques enable manufacturing companies to implement the complex processes necessary for revenue management in their business environments. Second, the high and further rising penetration of the modern society with semiconductor products empowers producers to enforce revenue management ideas, like dynamic pricing, in their business models. Third, the increasing pressure towards operational excellence has changed the mind-set of important decision-makers in the industry, which have traditionally been sceptical towards revenue management ideas. The semiconductor industry, and respectively industries consuming semiconductors, could be at the forefront of this move.

5 Increasing robustness and accuracy of demand fulfilment

This chapter bases on

Seitz, A., Grunow, M., 2017.
Increasing accuracy and robustness of order promises.
International Journal of Production Research, 55(3), 656-670.

Accurate order promising is a key requirement for customer satisfaction. Nevertheless, practitioners struggle with the reliability of the delivery dates they promise to customers. Consequently, the costs of demand fulfilment soar due to intensified communication, emergency processes in logistics and acquisition of costly external production resources.

Product and process flexibilities in supply chains that can be exploited in supply network planning are identified and formalized. Product flexibility is the possibility to produce several kinds of products from one predecessor product. Process flexibility is the possibility to use one production process to manufacture several products. In order to increase the accuracy and robustness of delivery dates, an order promising methodology able to deal with demand mix uncertainty and heterogeneous customer OLTs is developed. The approach anticipates changes in master production schedules made possible by product and process flexibilities.

A numerical study based on a case from the semiconductor industry demonstrates that the method increases the accuracy and robustness of order promises. For the studied case the consideration of process flexibility is more important for the generation of accurate and robust order promises than the consideration of product flexibility.

In the following, Section 5.1 introduces the problem of inaccurate order promises due to demand uncertainty caused changes in the master production schedule and illustrates the scientific contributions of the work presented in this chapter. Section 5.2 describes the planning environment. Section 5.3 and Section 5.4 explain the methodology and provide an illustrative example. Section 5.5 presents the framework and experimental design of the numerical study. The results are presented in Section 5.6, before Section 5.7 concludes the chapter.

5.1 Demand uncertainty caused changes of the master production schedule

Order promising as part of the order management of a company is known to be a key process for achieving long term business success (see e.g. Chen et al. 2002). Kilger and Meyr (2015) point out that customer retention and increase of market share strongly depend on the speed and reliability of the order promising process of a company. Furthermore, according to Oracle and Capgemini (2013) promising reliable delivery dates to the customer becomes more challenging. The complexity of global supply chains and the increase of order channels make companies face growing uncertainties in supply and demand. The study reveals that 42% of manufacturing and high-tech companies view accurately promising delivery dates as the main challenge in maintaining customer satisfaction. Additionally, Oracle and Capgemini (2013) show that inaccurate order promising is one of the main cost drivers for demand fulfilment since it causes additional efforts such as buying in costly external production resources in the short term, triggering emergency processes in logistics or intensifying communication with suppliers and customers in order to meet the promised delivery dates. Additionally, sales are lost because the employed order promising process is unable to anticipate possible changes in the master production schedule occurring after order arrival and therefore promises orders too late, i.e. after the earliest feasible delivery date. Thus, potential revenues are oftentimes not realised because customers cancel or do not place their orders in consequence of the late promise.

Changes of the master production schedule occur because the demand information used for supply network planning contains two types of uncertainty. First, demand forecasts are uncertain regarding volumes. Second, demand forecasts are uncertain regarding the proportion of the individual products in the demand mix. The sales department usually forecasts demands on the level of aggregate product families. However, production is planned on finished product level. Demand forecasts therefore have to be disaggregated for supply network planning. The rules used for this disaggregation usually build on ADI (see e.g. Hariharan and Zipkin 1995) and assumptions on how the forecast will realise on finished product level. The disaggregation obviously includes a second type of uncertainty, which is called demand mix uncertainty. It is defined as the uncertainty of the demand forecast with regards to the ratio of the individual product volumes, when the total demand of the product family is given. If realised orders deviate from the forecasted demand mix, the supply network planning process exploits flexibilities in the supply chain, i.e. product and process flexibilities, to change the master production schedule and to meet requested delivery dates of the customers. Mitigating the uncertainty regarding demand volumes on product family level is typically not within the scope of demand fulfilment approaches, but has to be dealt with in preceding processes. Therefore, the work focusses on demand mix uncertainty.

In order to increase the accuracy and robustness of order promises, supply chain flexibilities must be reflected in the order promising process. Otherwise the promised delivery date will be later than possible, increasing the risk for lost sales. In this chapter, a new method is presented that anticipates changes in the master production schedule using rules that consider product and process flexibilities of the supply chain. The approach increases the accuracy of order promises by reducing the amount of orders initially promised too late. At the same time, the approach increases the robustness of order promises by raising the amount of orders, whose initially promised delivery date is not changed throughout the demand fulfilment process.

The method is designed for supply chains with divergent material flows, flexible processes, and heterogeneous, i.e. varying and uncertain, customer OLT. Such environments are typical for the majority of industries. In divergent material flows several successor products can be produced from one predecessor product. The possibility to produce several kinds of products from one intermediate product or raw material is called product flexibility. The possibility to use one production process for the manufacturing of several products is called process flexibility.

Due to the heterogeneity of customer OLTs, companies typically plan and start their production on basis of a mix of orders and demand forecasts for finished products. The resulting master production schedule is used as a basis for the ATP process in real-time order promising. Newly arrived orders replace the respective demand forecast in the next run of the supply network planning process. In case the order deviates from the replaced forecast, the master production schedule is changed, possibly necessitating an update of the order promise. This update leads to reduced customer satisfaction because customers have to update their own master production schedules, possibly causing additional order changes that further reduce the robustness of the master production schedule.

The work has the following contributions:

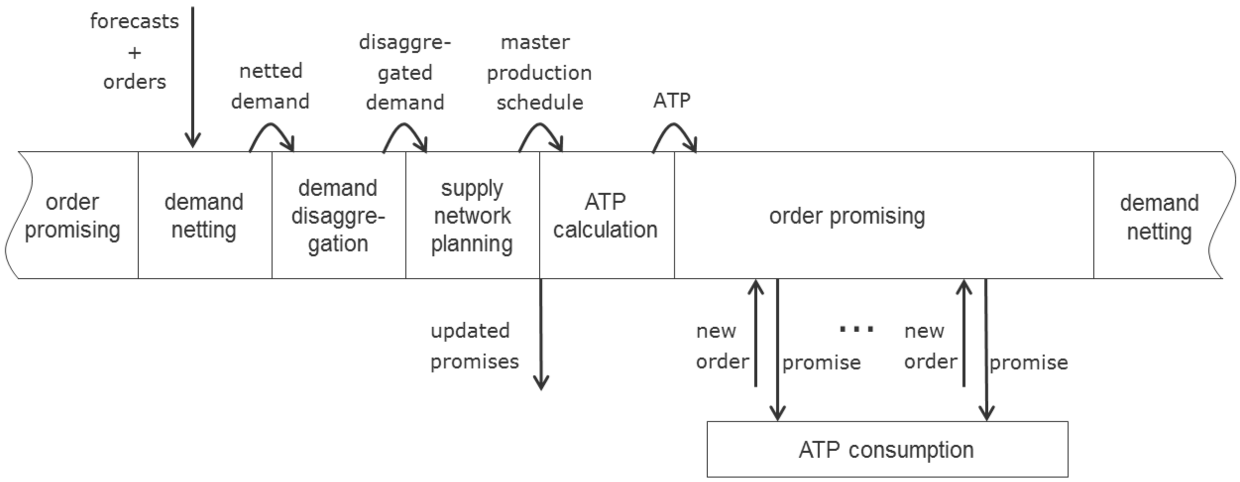

Figure 5: Planning process for order promising

- Product and process flexibilities in supply chains that can be exploited for supply network planning are identified and formalized.
- A new order promising methodology is presented, which is able to deal with demand mix uncertainty by considering supply chain flexibilities in ATP processes commonly used in practice. It is also capable of coping with heterogeneous customer OLTs.
- An increased accuracy and robustness of order promises is demonstrated in a numerical study of real-time order promising based on an industry case.

5.2 Planning process for order promising

Figure 5 shows the planning steps involved in the order fulfilment setting described in Section 5.1. The rolling process updating the master production schedule starts with the demand netting process. Here, a demand forecast is matched with the already realised customer orders. Forecast and orders differ in aggregation level. For ease of planning in the demand planning processes, forecasts are given on product family level, whereas orders are provided on finished product level. After demand netting, the netted demand is handed over to the demand disaggregation process, where the remaining forecasts are disaggregated to the finished product level.

Demand disaggregation is based on rules that forecast how the aggregate forecasts will realise on finished product level. This disaggregation of forecasts is subject to demand mix uncertainty. The resulting demand is used in a supply network planning process to generate the master production schedule. The process aims at fulfilling all demands on their requested delivery date. Afterwards, resulting delivery date changes are communicated to the customers in an updated order promise. Then, ATP information based on the master production schedule is calculated and forwarded to a real-time order promising process, which promises orders upon their arrival. These order promises might be updated after the following supply network planning, which exploits supply chain flexibilities described in Section 5.3 to fulfil orders on their requested delivery date. Finally, after every execution of the order promising step, ATP information needs to be updated in an ATP consumption process that reduces the available supply by the amount used to promise the new order.

For the development of a new approach that aims at increasing the accuracy and robustness of order promises, two requirements have to be considered. First, financial forecasting is based

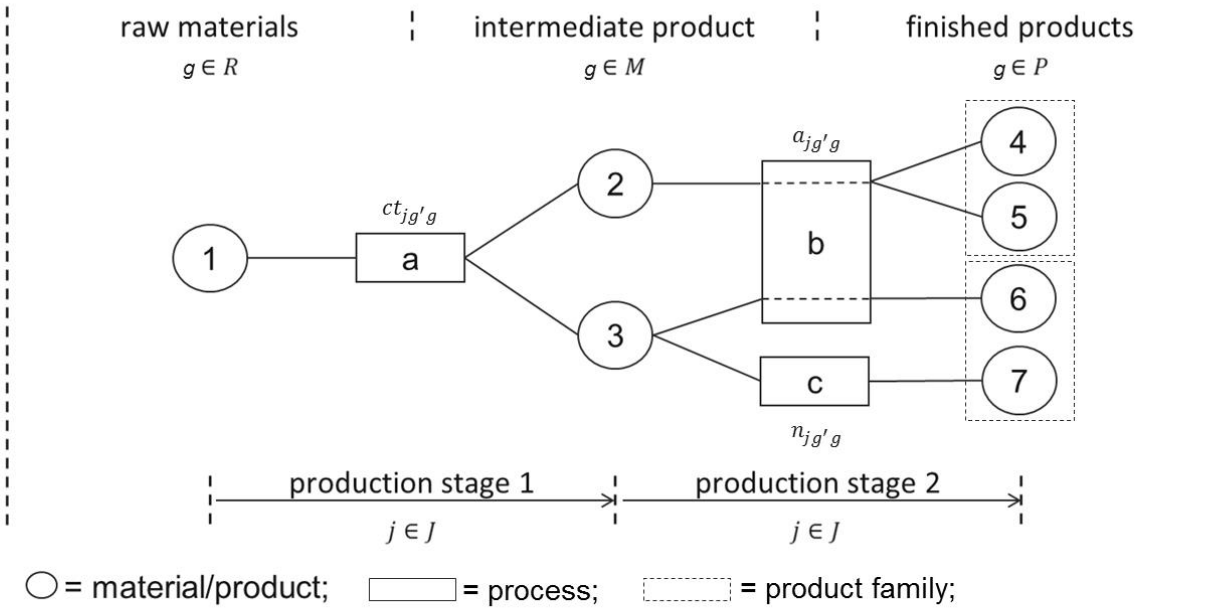

Figure 6: Formalization of supply chain flexibilities

on the current master production schedule. Since finished products differ significantly regarding generated revenues, production must be planned on that level. Second, existing order promising solutions of companies cannot be changed radically. Therefore, companies need new real-time order promising solutions that can easily be implemented into their process landscape. In most cases this means that order promising approaches based on complex capacity models cannot be applied in practice.

5.3 Representing supply chain flexibilities in ATP information

The order promising solution anticipates changes in the master production schedule after order arrival by representing supply chain flexibilities, i.e. process and product flexibilities, in ATP information. *Process flexibility* is defined as the possibility to use one production process $j \in J$ for the production of different intermediate products $g \in M$ or finished products $g \in P$. *Product flexibility* is the possibility to produce different finished products $g \in P$ or intermediate products $g \in M$ out of one intermediate product $g' \in M$ or raw material $g' \in R$. For better readability, existing combinations of j, g', and g are defined as $\theta = jg'g$, $\theta \in \Theta$. The production of g out of g' on j is characterized by the resource consumption factor a_θ and the BOM coefficient n_θ indicating the number of products g that are produced out of one unit of g' on j. In order to increase accuracy and robustness, the ATP calculations use the same cycle times ct_θ that are also used in supply network planning. Figure 6 illustrates the above described nomenclature on the example of a two stage production with divergent material flow. Note that the approach, however, is applicable to multi-stage production environments.

To represent product and process flexibilities in ATP information, an ATP cumulation step is inserted into the planning process shown in Figure 5. The calculation of the CATP is triggered every time a new order arrives. Figure 7 presents the resulting planning setup. Since the cumulation is based on the ATP information provided by supply network planning, it is easy to implement in existing demand fulfilment solutions.

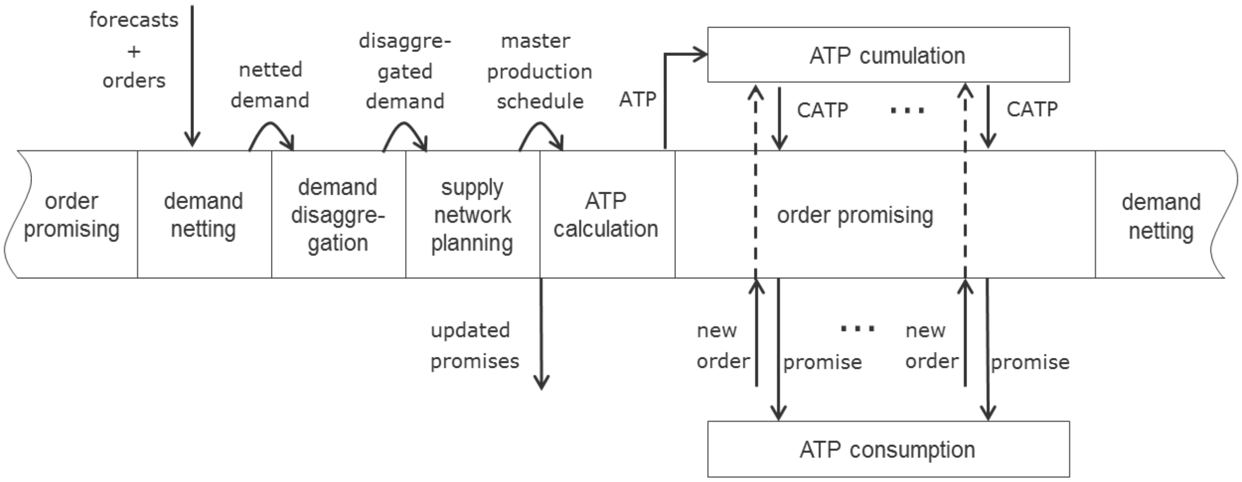

Figure 7: Planning process with ATP cumulation step

Representing supply chain flexibilities in ATP information entails several potential causes for errors. First, differences in capacity consumption factors of products, whose ATP quantities are cumulated, can cause the resulting CATP to be inaccurate, if the production capabilities of the shop floor are not represented accurately. Consequently, the order promises have to be changed after the next supply network planning run. This error is called resource consumption factor bias. Analogously, CATP quantities can be inaccurate, if the differences in the BOM coefficients of products, whose ATP is cumulated, are not reflected in the cumulation. This error is called BOM coefficient bias. Both aforementioned biases can be avoided by considering resource consumption factors and BOM coefficients in the ATP cumulation logic. Finally, ATP quantities on production rely on planned cycle times. In reality, however, the cycle times depend on the production schedule. The resulting error is called cycle time bias.

Four CATP types are formalized, which differ in the degree to which they take the above mentioned biases into account. Therefore, the planning horizon is structured into T periods of equal length. The ATP quantity atp_{gt} represents the amount of supply of finished product $g \in P$ becoming available in period $t \in T$. The sequence $S_{g'g}$ as is defined the sequence of processes $\langle j, j', \dots, j'' \rangle$ transforming a product $g' \in M \cup R$ into finished product g. Each $S_{g'g}$ produces $N_{g'g} = \prod_{j \in S_{g'g}} n_\theta$ units of g from one unit of g' with a cycle time $CT_{g'g} = \sum_{j \in S_{g'g}} ct_\theta$. G_{g^*t} is the set of finished products g, whose ATP quantities can be used to build CATP for product g^* in period t. To fulfil this condition, it first needs to be ensured that the supply chain has the flexibility to produce g^* instead of g from g'. Second, the sequences $S_{g'g}$ and $S_{g'g^*}$ must be equal. Third, it needs to be guaranteed that there is enough time to reroute intermediates $g'$from production of g to production of g^* without delay. For this, the period t, in which ATP is cumulated, must be at least $CT_{g'g}$ or $CT_{g'g^*}$ planning periods in the future. Hence, G_{g^*t} consists of all products g, whose production sequences $S_{g'g}$ equal the production sequence $S_{g'g^*}$ and for which $t \geq max(CT_{g'g^*}; CT_{g'g})$ holds. Furthermore, the parameters $\delta N_{g'gg^*} = (N_{g'g^*}/N_{g'g})$ and $\delta a_{g'gg^*} = \prod_{j \in S_{g'g}} min(1; a_{jg'g}/a_{jg'g^*})$ are defined for two sequences $S_{g'g}$ and $S_{g'g^*}$. Then, four types of CATP are formalized as follows:

SUM: $$catp_{g^*t}^{SUM} = \sum_{g \in G_{g^*t}} \left(atp_{gt} \right) \tag{39}$$

PROD: $$catp_{g^*t}^{PROD} = \sum_{g \in G_{g^*t}} \left(\delta N_{g'gg^*} \cdot atp_{gt} \right) \tag{40}$$

PROC: $$catp_{g^*t}^{PROC} = \sum_{g \in G_{g^*t}} \left(\delta a_{g'gg^*} \cdot atp_{gt} \right) \tag{41}$$

CUM: $$catp_{g^*t}^{COM} = \sum_{g \in G_{g^*t}} \left(\delta a_{g'gg^*} \cdot \delta N_{g'gg^*} \cdot atp_{gt} \right) \tag{42}$$

CATP type SUM (Equation (39)) sums up all ATP quantities atp_{gt} of the products contained in G_{g^*t} to build the CATP quantities $catp_{g^*t}^{SUM}$. SUM therefore does not take any of the mentioned biases into account. PROD (Equation (40)) eliminates the BOM coefficient bias in CATP quantities by multiplying the CATP quantities with the parameter $\delta N_{g'gg^*}$, which is the ratio of BOM coefficients of the sequences $S_{g'g}$ and $S_{g'g^*}$. PROC (Equation (41)) eliminates the consumption factor bias in CATP quantities by considering $\delta a_{g'gg^*}$, i.e. the ratio of consumption factor of the sequences $S_{g'g}$ and $S_{g'g^*}$. The minimum function in the calculation of $\delta a_{g'gg^*}$ ensures that CATP quantities are not falsely inflated, if the consumption factors of g^* are smaller than the ones of g. CUM (Equation (42)) combines the bias mitigation strategies of PROD and PROC. Note that the cycle time bias of CATP quantities cannot be addressed with the approach. However, it is suggested to install a feedback loop from the shop floor to communicate individual production starts to the order promising process in order to eliminate this bias.

5.4 Illustrating example

The differences of the CATP quantities resulting from the four CATP types are illustrated by using the example supply chain shown in Figure 6. Table 1 shows the values used for the parameters a_θ, n_θ, and ct_θ. The length of the planning horizon is set to $T = 6$ and the ATP quantities for all periods and products to a constant value of $atp_{gt} = 10$.

Table 2 presents the CATP quantities resulting from the different CATP types for an incoming order of product 5 ($= g^*$). The rows show for every period t (1st column) the sets G_{5t} (2nd column), containing all products that are used for the cumulation of ATP, and the resulting CATP quantities (3rd to 6th column).

In periods 0 and 1, no ATP can be cumulated ($G_{50} = G_{51} = \{5\}$), because the cycle time ct_{b25} required for transforming the intermediate product 2 through process b into the finished product 5 equals 2 (see Table 1). The parameters δN_{255} and δa_{255} equal 1. Therefore, in periods 0 and 1, all CATP quantities equal the ATP quantity of 10 for product 5 ($atp_{50} = atp_{51} = 10$).

In periods 2 and 3, the ATP quantities of the finished products 4 and 5 can be cumulated ($G_{52} = G_{53} = \{4; 5\}$), because the cycle times ct_{b24} and ct_{b25} are smaller or equal t. Before the CATP quantities can be calculated, the parameters δN_{245} and δa_{245} need to be determined (δN_{255} and δa_{255} are already known from above). Using the equations presented in Section 5.3 and the parameters shown in Table 1, δN_{245} and δa_{245} are calculated as follows:

$$\delta N_{245} = \frac{N_{25}}{N_{24}} = \frac{n_{25}}{n_{24}} = \frac{4}{2} = 2$$
$$\delta a_{245} = min\left(1; \frac{a_{24}}{a_{25}}\right) = min\left(1; \frac{1}{2}\right) = \frac{1}{2}.$$

Table 1: Example CATP types: parameters

		$\theta = jg'g$					
		$a12$	$a13$	$b24$	$b25$	$b36$	$c37$
Para-meter	a_θ	1	2	1	2	2	1
	n_θ	1	2	2	4	1	1
	ct_θ	2	1	1	2	2	3

The CATP quantities are then determined by Equations (39) to (42). They result in:

$$catp_{52}^{SUM} = catp_{53}^{SUM} = atp_{42} + atp_{52} = 10 + 10 = 20.$$
$$catp_{52}^{PROD} = catp_{53}^{PROD} = \delta N_{245} \cdot atp_{42} + \delta N_{255} \cdot atp_{52} = 2 \cdot 10 + 1 \cdot 10 = 30.$$
$$catp_{52}^{PROC} = catp_{53}^{PROC} = \delta a_{245} \cdot atp_{42} + \delta a_{255} \cdot atp_{52} = \frac{1}{2} \cdot 10 + 1 \cdot 10 = 15.$$
$$catp_{52}^{COM} = catp_{53}^{COM} = \delta N_{245} \cdot \delta a_{245} \cdot atp_{42} + \delta N_{255} \cdot \delta a_{255} \cdot atp_{52} =$$
$$= 2 \cdot \frac{1}{2} \cdot 10 + 1 \cdot 1 \cdot 10 = 20.$$

Analogously, in periods 4 and 5, the CATP quantities are composed of ATP supply from the products 4, 5, and 6 ($G_{54} = G_{55} = \{4; 5; 6\}$). This is because these products can all be produced from raw material 1 on the process sequence $S_{14} = S_{15} = S_{16} = \langle a, b \rangle$ and the respective cycle times $CT_{14} = ct_{a12} + ct_{b24} = 3$, $CT_{15} = 4$, and $CT_{16} = 3$ are smaller or equal t. Using the equations presented in Section 5.3 results in the parameter values $\delta N_{145} = 2$, $\delta N_{155} = 1$, $\delta N_{165} = 2$, $\delta a_{145} = 0.5$, $\delta a_{155} = 1$, and $\delta a_{165} = 1$, which lead to the CATP quantities presented in Table 2, when Equations (39) to (42) are applied.

As can be seen in Table 2, all CATP types result in different CATP quantities. BOM coefficients and resource consumption factors are considered to a different extent. The type SUM simply sums up all ATP quantities in K_{5t}. This type is therefore most suitable for environments, in which all products and materials have equal or at least similar BOM coefficients and consumption factors.

The CATP type PROD considers differences in BOM coefficients between products used in the cumulation. This CATP type is therefore most suitable for environments with large differences in BOM coefficients between products, but small or no differences in resource consumption. In the example, PROD results in the highest CATP quantities because the BOM coefficients of product 5 are larger than those of products 4 and 6. As more units of product 5 than of the products 4 and 6 can be produced from the same amount of intermediate products or raw materials the CATP quantities are increased accordingly.

CATP quantities of the type PROC consider differences in resource consumption factors between products used in the cumulation. This CATP type is most suitable for environments with large differences in resource consumption between products, but small or no differences in BOM coefficients. In the example, PROC results in small CATP quantities because the consumption factors of product 5 are higher than those of products 4 and 6.

Finally, the CATP type CUM considers both resource consumption and BOM coefficients. It is most suitable for environments with large differences between products in both factors. In the example, CUM results in values for the CATP quantities, which lie between those generated by

Table 2: Example CATP types: resulting CATP quantities

t	G_{5t}	$catp_{5t}^{SUM}$	$catp_{5t}^{PROD}$	$catp_{5t}^{PROC}$	$catp_{5t}^{CUM}$
0	5	10	10	10	10
1	5	10	10	10	10
2	4;5	20	30	15	20
3	4;5	20	30	15	20
4	4;5;6	30	50	25	40
5	4;5;6	30	50	25	40

PROD and PROC. The high BOM coefficient of product 5 increases the CATP while the high resource consumption factor of product 5 decreases the CATP.

5.5 Numerical study

The advantages of the approach are illustrated in a case study from the semiconductor manufacturing industry. Supply chains in this industry show divergent material flows. Silicon wafers are transformed into dozens of different variants of integrated circuits. It also shows flexible processes. The machines in this industry are highly capital intensive, leading to the necessity of providing the flexibility to process different products on the same resources. Hence, supply chains in this industry show the product and process flexibilities exploited by the approach. Furthermore, the production cycle times of four to six months are typically much longer than customer OLTs. Therefore, production is started based on aggregate demand forecasts, which, due to this long horizon, are subject to significant demand mix uncertainty. For a more detailed description of the characteristics of semiconductor manufacturing, the interested reader is referred to Mönch et al. (2013).

The method is compared to conventional order promising described in Section 5.2 and a CTP approach that uses the supply network planning process for real-time order promising. Therefore, first an overview of the used framework for the numerical study is given in Section 5.5.1. Afterwards, the design of experiments is discussed in Section 5.5.2.

5.5.1 Framework

For the numerical study, the planning processes presented in Figure 7 are implemented. Every time unit, aggregated demand forecasts are netted with already realised orders from earlier periods by subtracting the order quantity from the forecast quantity in the period of the requested delivery date. Afterwards, the remaining forecast quantities are disaggregated. Here, the rule employed at the case company is implemented, where the historical proportions of demands on finished product level are used to generate the demand mix.

For supply network planning the approach presented in Section 2.2.2 is used. Promises of not yet delivered orders are updated, if the master production schedule allows earlier delivery or requires a postponement. After that, the production completion times of products are used to generate the ATP information.

After supply network planning, several orders from different customers realise. On order arrival, ATP is cumulated (Section 5.3) and order promising is executed. The model used for order promising is an adaptation of the approach shown in Section 2.3.3. The model, which is modified to allow order promising based on CATP quantities and promise orders based on virtual per-unit profits $prof_t$ of fulfilling an order in period t, decides on the CATP quantities consumed to promise an incoming customer order o. Equations (43) to (46) describe the model.

Maximise

$$z = \sum_t prof_t c_t, \tag{43}$$

subject to

$$\sum_t c_t \leq q_o, \tag{44}$$

$$c_t \leq catp_t, \qquad \forall t \in T; \tag{45}$$

$$c_t \geq 0, \qquad \forall t \in T. \tag{46}$$

The objective function (43) maximises the profit generated by fulfilling o. To set the per-unit profits $prof_t$, Equations (47) and (48) are used, which make sure that on-time fulfilment of an order is most preferable and early fulfilment is preferred over late fulfilment. The parameters $prof_0$, q_o, and t_o are defined as a base profit, the ordered quantity, and the requested delivery period of o, respectively.

$$prof_t = \frac{prof_0}{q_o}(T - t_o + t), \qquad \forall t \in T | t \leq t_o. \tag{47}$$

$$prof_t = \frac{prof_0}{q_o}(T - t) \qquad \forall t \in T | t > t_o. \tag{48}$$

Constraints (44) ensure that the consumed CATP quantities do not exceed the ordered quantity. Constraints (45) state that the consumed CATP quantities must not exceed available CATP, which is provided by the ATP cumulation step described in Section 5.3. The conventional order promising approach described in Section 5.2 does not use the ATP cumulation step but promises orders based on finished product ATP of the requested product alone. In this case, the parameters $catp_t$ in Constraints (45) equal the ATP quantities derived from the master production schedule. Constraints (46) are non-negativity constraints.

After order promising, the promised delivery quantities are determined using Equations (29) and (30) from Section 2.3.3. Finally, the ATP information is updated in an ATP consumption step. In the CTP approach, the supply network planning model is used to generate the real-time order promises.

As performance indicators the robustness, late and early promised orders are used, which are the share of orders, whose initially promised delivery date equals, is later or earlier than their delivery date, respectively. A decrease in the share of late promised orders represents an increase of the accuracy of order promises. The indicators early promises and late promises are defined in Equations (49) and (50), in which O is the set of all customer orders that realised over the horizon T^s of the numerical study, t_o^p is the promised delivery period for order o, t_o^d is the realised delivery period for order o, and dq_{ot} is the delivered quantity for order o in period t. The values of the indicators robustness and accuracy are derived as defined by Equations (33) and (34) in Section 2.3.4.

$$early\ promises = \frac{\sum_{o \in O, t \in T^s | t_o^p < t_o^d} dq_{ot}}{\sum_{o \in O} q_o} \tag{49}$$

$$late\ promises = \frac{\sum_{o \in O, t \in T^S | t_o^p > t_o^d} dq_{ot}}{\sum_{o \in O} q_o} \tag{50}$$

To be able to investigate the pure effects of demand mix uncertainty on the performance of the order promising approaches and eliminate all other sources of uncertainty typically appearing in a real world environment, the following assumptions are made:

- Capacities and cycle times of processes are fixed, constant and deterministic.
- No buffer stocks are considered.
- There is no restriction on the availability of raw materials.
- Customers do no cancel or reschedule orders.
- Aggregated demand forecasts do not contain a forecast error.

5.5.2 Experimental design

The semiconductor industry shows divergent material flows and heterogeneous customer OLTs. The raw materials are silicon wafers, while intermediate products are processed wafers or separated unfinished chips and finished products are the finished chips. The processes modelled in supply network planning are called bottlenecks. These are representations of machine groups which are typically constraining the capacity of a semiconductor supply chain. Two diode production lines of a large European company are investigated. The products have process cycle times between five and eight weeks.

Each cumulation logic is tested for every combination of levels of the factors supply chain flexibility, demand mix uncertainty, and customer OLT heterogeneity. For the factor cumulation logic, the four CATP types presented in Section 5.3 are implemented, the conventional ATP approach described in Section 5.2 and the CTP approach mentioned in Section 5.5.1.

For the factor supply chain flexibility, two product lines of the case company are investigated, which show low and high supply chain flexibility. To measure the supply chain flexibility of a product line an indicator defined in Chatterjee et al. (1984) as the average number of alternative parts that can be manufactured on a production sequence is used. The investigated product lines show flexibility values of 1.5 (low) and 4.5 (high). The capacity of each process is chosen such that the expected capacity utilization of each process equals 80%, which is the value used at the case company to plan production.

To represent demand mix uncertainty, first, two orders with a random customer OLT for each product and period are generated. In order to generate low, medium, or high demand mix uncertainty, one product within each product family is chosen randomly, to which the total demand of one, two, or three randomly chosen products of the same product family is assigned. To choose products, a uniform distribution is used. The resulting demand mix uncertainty is measured in terms of the SMAPE. The described approach results in SMAPE values of 30% (low), 35% (medium), and 43% (high) for Product Line I and 35%, 52%, and 69%, for Product Line II. Note that, even though the order stream is generated before each run, customer orders are not known to the demand fulfilment processes until they realise.

For the factor customer OLT heterogeneity, the levels MTO-skewed, ATO-skewed, MTS-skewed, and uniform are investigated. Note that customer OLTs remain heterogeneous for all levels. In the MTO-skewed case, the majority of orders arrives with an OLT greater than the production cycle time. This case represents businesses, in which the majority of the products produced are highly customer specific and costumers provide long term demand forecasts. For

Table 3: Design of experiments for CATP methodology

Factor	Levels	Count
Cumulation logic	conventional, SUM, PROD, PROC, CUM, CTP	5
Supply chain flexibility	low flexibility, high flexibility	2
Demand mix uncertainty	low, medium, high	3
customer OLT heterogeneity	MTO-skewed, ATO-skewed, MTS-skewed, uniform	4
Replications		10
		1200

the level ATO-skewed, the majority of orders arrives before the final assembly step. Here, the majority of the products are not customer specific during production but in product assembly, so that customers place their orders with a customer OLT longer than the assembly cycle time. The level MTS-skewed stands for environments, in which the majority of products are not customer specific and orders arrive with short customer OLTs. In the perfectly heterogeneous case, orders arrive with a uniformly distributed customer OLT. A planning horizon length of 13 weeks is used and the maximum and minimum customer OLTs are set to 13 and 0 weeks, respectively.

Ten replications for each combination of factors are generated and the replication length is set to 52 weeks. Table 3 summarizes the described design of experiments. The material flow diagrams of the investigated supply chains as well as graphical representations of the used customer OLT profiles can be found in Appendix B.

5.6 Results

The numerical results presented in the following were derived on a personal computer with a Windows 7 (32 bit) operating system with 4 GB memory and an Intel® Xeon® CPU with 2.53 GHz. The study was implemented in Java using the additional library Stochastic Simulation for Java for implementing probability distributions and the CPLEX Java API for solving the supply network planning and order promising methods. All figures show the performance relative to the conventional ATP approach, which is described in Section 5.2.

Figure 8 presents an overview of the relative performance of the CATP approach compared to conventional ATP. The graphs show that CATP outperforms conventional ATP in terms of late promises, i.e. accuracy, and robustness. However, CATP increases early promises, but the effect is less strong than the reduction of late promises. In terms of robustness, the CATP types PROC and CUM perform worst amongst the CATP approaches, because they reduce the amount of available CATP in case of varying resource consumption factors in the production sequence of cumulated products. However, this also leads to a distinct positive effect on early promised orders, because of the elimination of the resource consumption factor bias. Since the performances of the CATP types SUM and PROD do not significantly differ, it is concluded that the elimination of the BOM coefficient bias alone does not have strong influence on CATP performance for the investigated environment. This is because in semiconductor supply chains products of the same family typically have similar BOM coefficients. The CTP approach outperforms the best CATP approach by more than 20%. The reason for this is that orders are promised with complete information about process utilization, work in progress and remaining

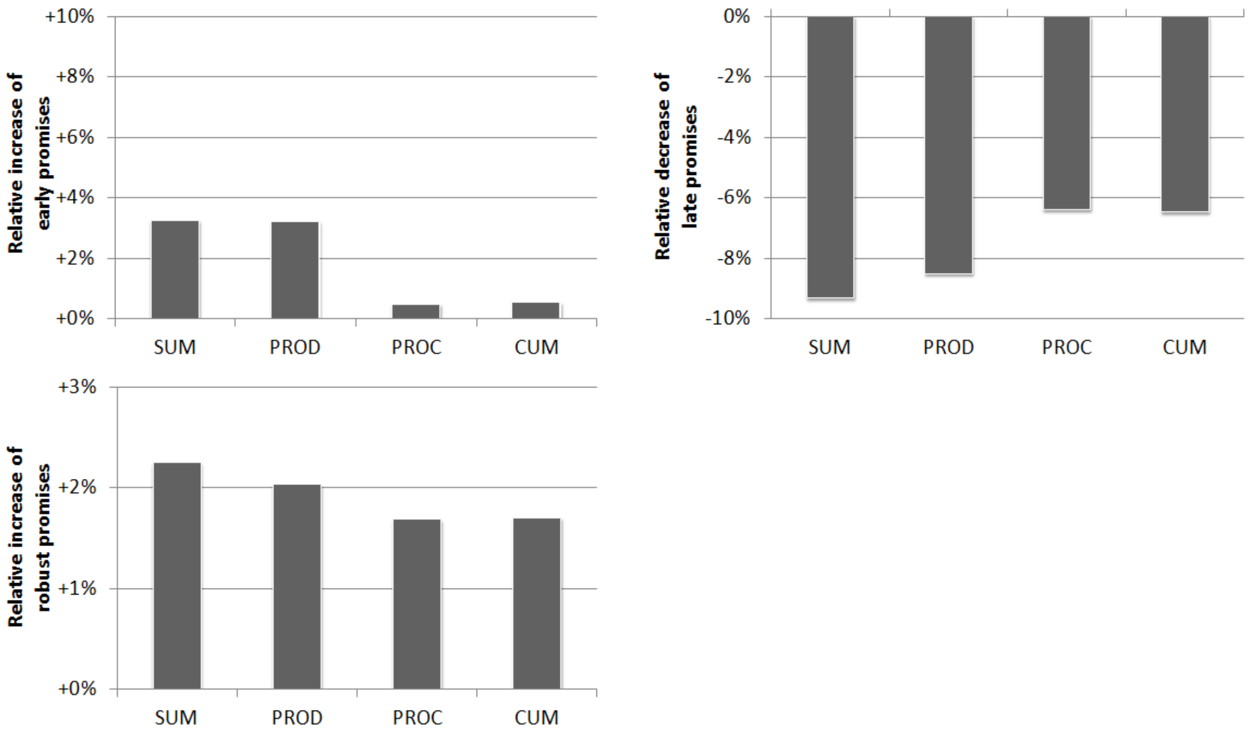

Figure 8: Relative performance overview CATP vs. conventional ATP

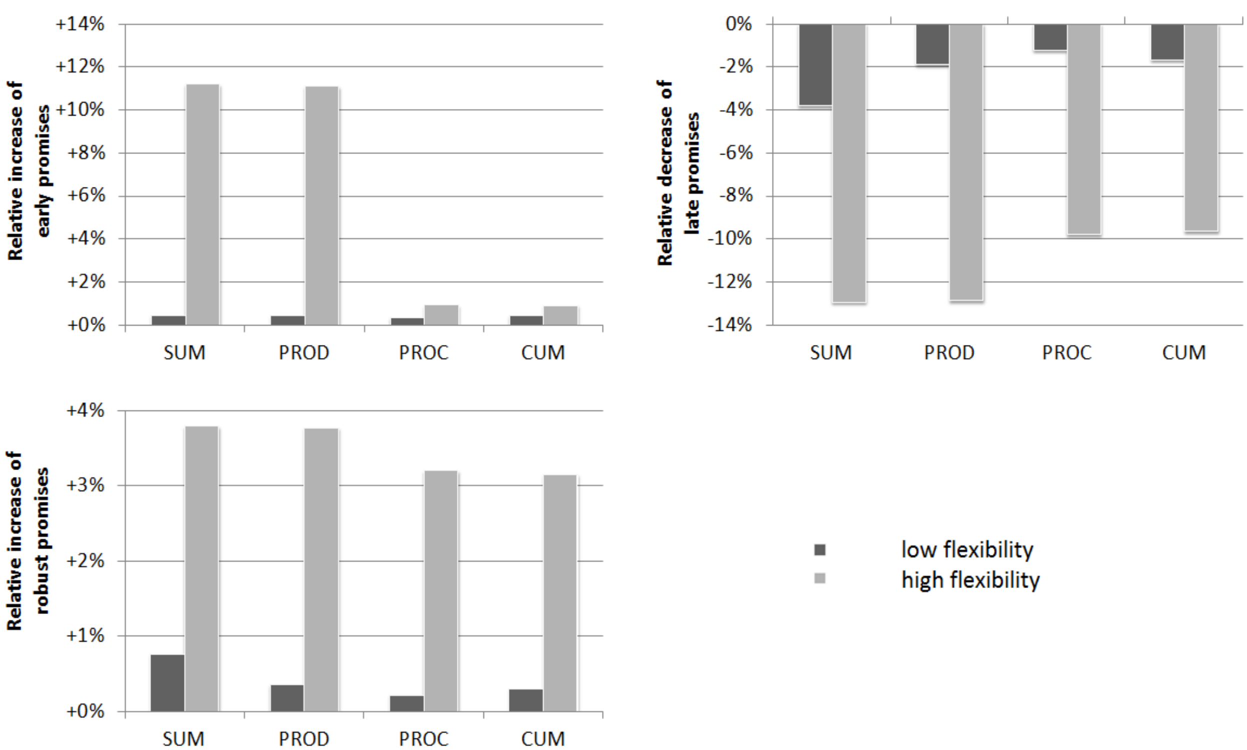

Figure 9: Relative performance of CATP approaches depending on supply chain flexibility

flexibilities in the supply chain. ATP and CATP solutions, on the other hand, cannot make use of, e.g., free capacities and production start information or reschedule planned production.

However, CTP approaches usually cannot be applied in practice because they are computationally expensive and hard to implement into existing order promising solutions. For better visibility of the performance of the CATP approaches, the performance of CTP is not shown in Figure 8 and the remainder.

Figure 9 shows the relative performance of the CATP approaches depending on the supply chain flexibility. It can be seen that their advantage in terms of accuracy, i.e. the decrease of late promises, and robustness grows substantially with increasing supply chain flexibility. However, the amount of early promises increases significantly for the CATP types SUM and PROD because of the resource consumption factor bias contained in their CATP quantities.

The relative performance of CATP approaches for different demand mix uncertainties is shown in Figure 10. The advantage of CATP approaches in terms of accuracy and robustness decreases with increasing demand mix uncertainty. Interestingly, also the increase of early promises decreases with increasing demand mix uncertainty for the CATP types PROC and CUM while it increases for SUM and PROD. For problem instances with high demand mix uncertainty, PROC and CUM do not show any significant advantage over the conventional ATP approach. However, the SMAPE value for these scenarios is above 50%, which is significantly higher than in most real world environments. It is therefore concluded that CATP leads to an improvement of order promising accuracy and robustness in practice.

Figure 11 shows the relative performance of the CATP approach for different customer OLT profiles. Since the above analysis showed that the performance of the CATP types SUM and CUM is representative for the performance of the types PROD and PROC, respectively, only the performance of SUM and CUM are shown for better visibility. Both types perform best in the ATO-skewed problem instances because in both product lines investigated all products whose ATP can be cumulated differ in their consumption factors only in one specific production step. Since in the ATO-skewed customer OLT profile the majority of orders arrives after this step, the CATP quantities used for order promising do not contain a consumption factor bias. Additionally, orders arrive with a customer OLT long enough to prevent a cycle time bias in CATP quantities. Furthermore, the advantage of CATP is higher for MTO-skewed environments than for MTS-skewed environments. This is intuitive since when customer OLTs are long, more flexibilities can be exploited to meet the promised delivery dates.

5.7 Conclusion

Accurately promising delivery dates for customer orders is indispensable for maintaining customer satisfaction. However, in global supply chains it is also one of the biggest challenges. In this chapter, a new order promising method called CATP is presented. It is designed for supply chains with divergent material flows, flexible processes, and heterogeneous, i.e. varying and uncertain, customer OLTs. In such environments order promises are often given in real-time and production is planned in regular intervals on basis of forecasts subject to demand mix uncertainty and realised customer orders.

CATP anticipates changes in the production plan due to newly arrived orders and increases the accuracy and robustness of order promises. The anticipation is done by representing supply chain flexibilities in ATP information used for real-time order promising. The flexibilities considered are product and process flexibility, which are identified and formalized. The method

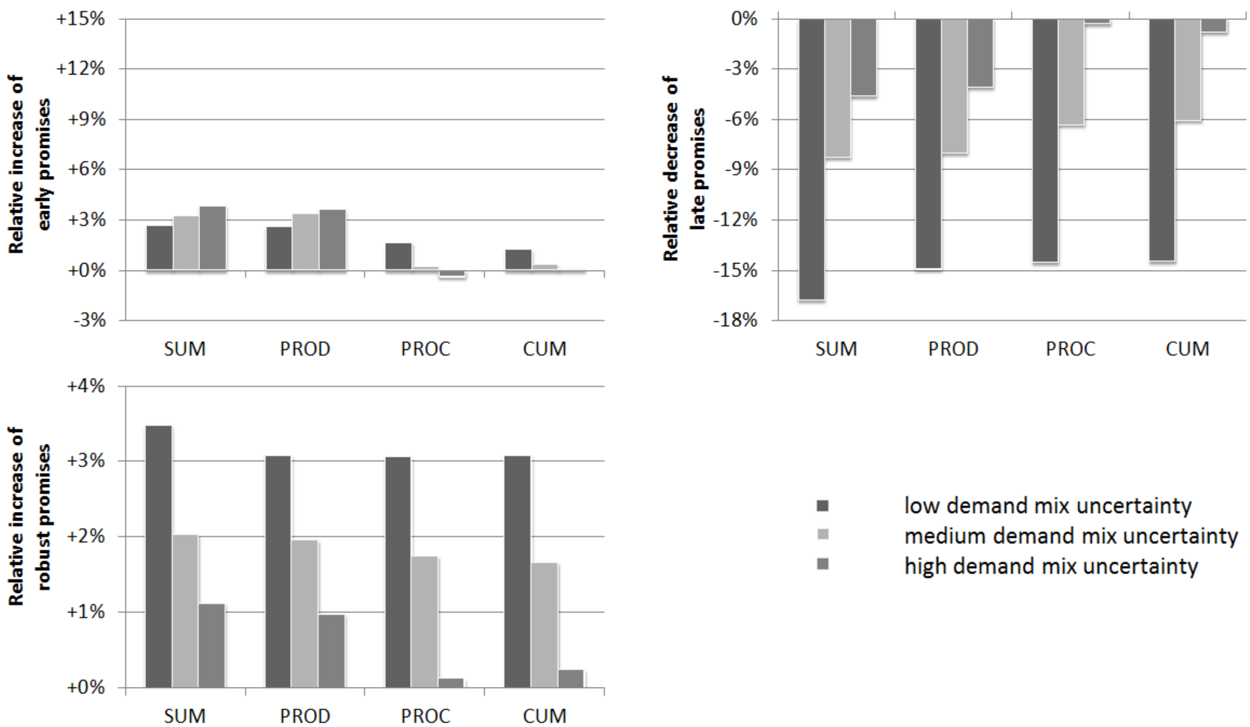

Figure 10: Performance of CATP approaches depending on demand mix uncertainty

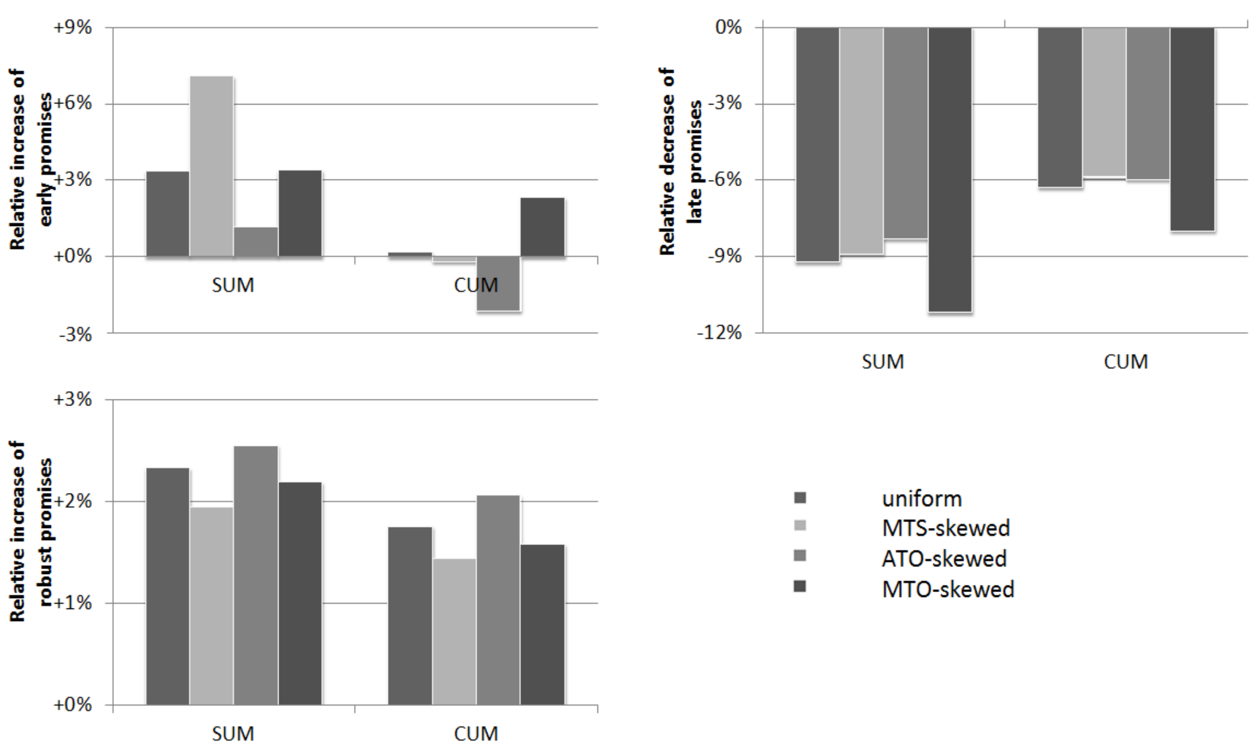

Figure 11: Performance of CATP approaches depending on order lead time

cumulates finished product ATP information based on the product individual characteristics production cycle time, resource consumption factor and BOM coefficient.

In a numerical study based on a case from the semiconductor industry, the performance of CATP is compared to a conventional approach in environments with different degrees of supply chain flexibility, demand mix uncertainty, and customer OLT heterogeneity. For benchmarking reasons, the method is also compared to a CTP solution which uses the production planning process to promise incoming orders in real-time and therefore poses an upper bound for the performance of CATP and conventional order promising. However, because of its high computational effort and necessity of substantial changes in planning processes commonly employed in practice, CTP cannot be applied in most supply chains.

The study shows that CATP increases the accuracy and robustness of order promises by reducing the amount of orders receiving an initial order promise later than the earliest feasible delivery date. Thus, the approach increases the order winning probability for a company since the customers more often receive the earliest possible delivery date for their order. It is also shown that considering resource consumption factors and BOM coefficients in ATP cumulation reduces the amount of orders initially getting promised earlier than feasible. The advantage of the CATP approach grows with increasing customer OLTs and degree of flexibility in the supply chain. Only if the demand mix uncertainty is high, CATP does not show any advantage over conventional order promising. This effect, however, can only be seen for unrealistically high values of forecast errors. Finally, it is found that for the studied case the consideration of process flexibility is of more value for the generation of accurate and robust order promises than product flexibility.

The research in this chapter focusses on representing the supply chain flexibilities in order promising that are exploited by production planning. Consequently, it is assumed that there are no other sources of uncertainty except demand mix uncertainty. Interesting directions for future research are to study the dependency of the robustness of given real-time order promises regarding the finally realised delivery date on the accuracy of estimated cycle times, especially when also demand forecast volumes are uncertain and these uncertainties interact with demand mix uncertainty and safety stocks. In such scenarios, also unknown future demand needs to be accounted for in order promising, especially when customers are heterogeneous regarding demand fulfillment relevant characteristics (e.g. profitability or forecast accuracy). In industrial environments, this is typically done by means of supply allocation methodologies, which should be extended to also consider forecast accuracy. Such approaches are presented in Chapters 6 and 7. When utilization is fluctuating and cycle times are dynamic, the CATP method needs to be extended to reflect additional information about the shop floor. Here, the consideration of order individual production cycle times as well as supply chain capabilities, i.e. unused capacities and swopping possibilities in the production schedule, are potential starting points.

The approach aims at increasing the accuracy and robustness of order promises to ensure customer satisfaction. To investigate an immediate impact on profitability, research is required that integrates the approach with methodologies from revenue management.

6 Considering the bias of advance demand information in allocation planning

This chapter bases on

Seitz, A., Grunow, M., Akkerman, R., 2016b.

Data Driven Supply Allocation to Individual Customers Considering Forecast Bias.

Available at SSRN: https://ssrn.com/abstract=2813835.

In this chapter, a data driven allocation planning (DDAP) approach is proposed, which exploits increasingly available data on individual customers and products by allocating supply on a highly granular level at high planning frequencies. The method considers the demand forecast bias of customers, supports an efficient supply allocation and incentivises the customers to communicate truthful forecasts. Using the approach in a numerical study based on the semiconductor industry, it is demonstrated that the approach increases overall service levels, especially for customers with truthful forecasts, and reduces excess allocations, leading to lower inventory levels. The analysis further shows that the allocation efficiency increases with the granularity level and the predictive quality of the available data.

This chapter is organized as follows. In Section 6.1 the problem of inefficient supply allocations because of rationing gaming of customers is introduced and the contributions of the work presented in this chapter are illustrated. Section 6.2 explains the allocation planning approach in detail. Section 6.3 introduces a case from the semiconductor industry, describes the experimental design of the numerical study, and presents its results. Conclusions are presented in Section 6.5.

6.1 Big data enables companies to reduce the risk of inefficient supply allocations

Shortening economic and product life cycles lead to increasing demand variations, which are amplified through the supply chain by the so-called bullwhip effect. For example, the growth rates of the semiconductor market (without memory chips), whose companies are typically located upstream in the supply chain, varied between −40 % and +50 % since the beginning of 2009 (WSTS Inc. 2015). Consequently, periods of supply shortage occur more frequently. Hence, allocation planning, i.e. deciding on when and how to fulfil which customer's demand, gains importance.

AP is part of a demand fulfilment process, typically implemented in software systems such as advanced planning systems (see Figure 12). It reserves quantities of inventory and planned supply receipts, together termed ATP, for certain customer segments. These supply reservations, called AATP, are then used to confirm delivery dates for incoming customer orders in a real-time order promising step. The communication with the customer is often fully automated and occurs at high frequencies, allowing the allocation planning to also be performed at high frequencies.

With the recent advances in big data tools, companies are able to monitor the ordering behaviour of their customers on the granularity level of individual customers and final products. The higher transparency of the customers' ordering behaviour provides opportunities to increase the efficiency of supply allocation. However, conventional allocation planning approa-

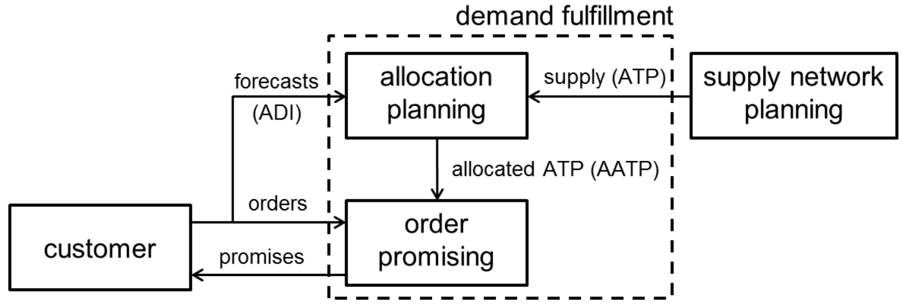

Figure 12: General structure of a demand fulfilment process

ches based on segmentation are not able to exploit the available data because customers are typically clustered according to profitability and not according to their ordering behaviour.

From the literature it is known that in order to satisfy their demands, the customers strategically inflate their forecasts in supply shortage situations to game the allocation planning procedure of their supplier. Intuitively, one would assume that customers inflate their orders by a fixed bias. However, the data from a large European semiconductor manufacturer, which is used in a case study, shows that the consumers usually communicate a stable forecast, but order only smaller volumes at irregular intervals. Figure 13 shows the demand errors and the forecast bias resulting from typical customer forecasting and ordering behaviour. The error is defined as the part of the forecast that is not translated into actual orders relative to the total forecast. Its strategic component, i.e. its non-random part, is defined as the forecast bias. In Section 6.2.1, the exact definitions of the forecast error and the forecast bias are provided.

For a supplier, such customer behaviour, commonly referred to as rationing gaming (see e.g. Lee et al. 2004), leads to the risk of inefficient supply allocation. Stocks are created for forecasts, which are not consumed by subsequent orders. This results in high storage costs and, more importantly, low overall service levels due to the limited supply.

In order to counteract the negative consequences of the rationing game, new approaches are needed to incentivise customers to truthfully forecast their demand. Herein, the results of big data tools monitoring historical demand forecast biases can be exploited to identify systematic gaming behaviour.

This paper develops an allocation planning approach that allocates ATP supply to individual customers. The approach is developed for industrial environments in which standardised goods are mass produced and businesses use a make-to-stock strategy in supply planning. Using a notification-release order cycle, available supply is first allocated on the basis of demand forecasts, i.e. notifications, provided by the customers and released later upon order reception. The approach presented here considers customer demand forecasts and their historical biases as well as their profitability. In particular, the following contributions are made:

- A new allocation planning approach is presented, which exploits increasingly available data on individual customers and products by allocating supply on a highly granular level at high planning frequencies. More specifically, the methodology examines the demand forecast bias of customers and thereby supports efficient supply allocation and incentivises the customers to communicate truthful forecasts.
- Using the approach in a numerical study based on the semiconductor industry, it is demonstrated that

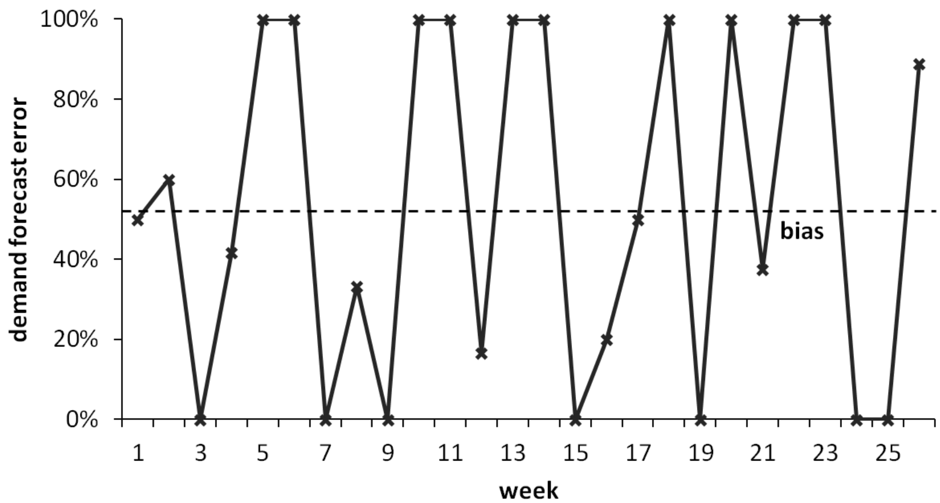

Figure 13: Demand forecast bias graph for a typical customer demand pattern

- the proposed methodology leads to lower average stock levels and an increased overall service level, especially for customers with truthful forecasts,
- the allocation on individual customer level is only valuable when additional data such as forecast bias and lead time is exploited on this granularity level.

6.2 Data driven allocation planning methodology

Figure 14 gives an overview of the DDAP methodology. It is developed for the single-product case and divided into a mid-term and two short-term planning processes, executed in a rolling horizon fashion. The approach can also be applied to the multi-product case provided there is no substitution between products. The mid-term customer PAS determination process, described in Section 6.2.1, uses the profitability and historical forecast accuracy of the customers to determine their individual profitability accuracy score (PAS_i). The resulting score of individual customers is used in the decisions of the short-term processes allocation planning and order promising.

The short-term allocation planning method, which is developed in Section 6.2.2, reserves supply quantities becoming available in planning period $t \in T$ for forecasted demand being due in period $\tau \in T$. It fulfils the demands in the sequence of the PAS of the customers. The allocation planning step is part of the short-term planning of the company. It is so because industrial customers update their demand forecasts in a frequent, short-term manner (e.g. every day). Accordingly, supply allocations are updated with a high frequency in order to make use of the newly available data.

The real-time order promising approach employed in the DDAP methodology is an adaptation of the model presented in Section 2.3.3. Because DDAP allocated supply to individual customers $i \in I$, the set of customer segments K equals the set of customers I. Hence, K is replaced with I, K_{i^*} with I_{i^*}, k with i and k^* with i^* in Equations (23) to (30) for the adaptation. Furthermore, $score_k^{seg}$ is replaced with PAS_i. Note that customers do not have to place an order in every time period and there is no predetermined sequence in which the customers order.

Finally, the advantages of the allocation planning model over conventional allocation planning (CAP) are illustrated in Section 6.2.3.

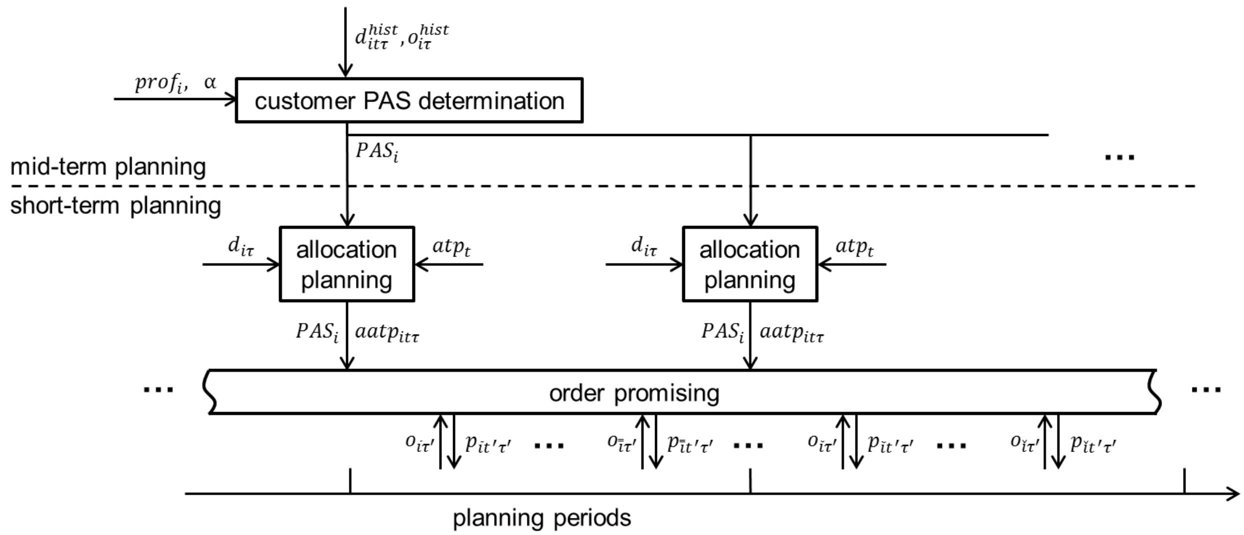

Figure 14: Rolling horizon scheme for data driven allocation planning

6.2.1 Mid-term customer PAS determination

For the DDAP approach, the concepts of the forecast error and the forecast bias have to be distinguished. While the forecast error is simply defined as the deviation of the customer forecast from the final order, the forecast bias denotes the systematic or strategic deviation. Hence, the forecast error consists of a random component and a systematic or strategic forecast bias. If the forecast bias of a customer is positive, they systematically or strategically inflate their demand forecasts, i.e. show rationing gaming behaviour. The DDAP approach identifies rationing gaming of the customers and considers it in allocation planning in order to incentivize customers to truthfully forecast their demands.

As described in Section 6.1, industrial customers displaying rationing gaming behaviour usually do not inflate their demand forecasts by a fixed ratio. Instead, constant demand forecasts of the (anticipated) maximum possible order size are given; the actual order sizes then experience substantial volatility (see Figure 13). Hence, an allocation approach will not lead to an efficient supply allocation if it integrates demand forecast bias information by discounting demand forecasts by a certain predetermined constant that represents the strategic forecast inflation of the customer. However, it is still possible to identify strategic gaming behaviour of the customer by determining the statistically significant positive deviation of the customers' forecasts from their final orders.

The DDAP approach determines customer scores PAS_i , which are composed of the historical forecast accuracy acc_i and profitability $prof_i$ of the customers $i \in I$. They represent the priority of a customer for the supply allocation decision. Note that, here, acc_i is defined as the degree to which the forecast of customer i is free of a strategic forecast bias. The measure does not take the random component of the forecast error into account.

A factor $\alpha \in [0; 1]$ represents the decision maker's trade-off between forecast accuracy acc_i and profitability $prof_i$ in the determination of PAS_i . At an α-value of 0 and 1, PAS_i is determined solely by $prof_i$ and acc_i, respectively.

The customer scores are determined using a four-step data analysis, which is executed in a mid-term rolling horizon fashion, e.g. once every year. For this analysis, the customer profitability $prof_i$, the historical demand forecast $f_{it\tau}^{hist}$ provided by customer i in period t for period τ, and the history of orders $q_{i\tau}^{hist}$ placed by customer i with a due-date in period τ are analysed.

The DDAP approach is designed for large industrial suppliers, which usually only serve large customers with a significant revenue contribution directly. Other, smaller customers are served through distributors, who cumulate the demands of numerous small customers and thus become large customers for the supplier as well. In such environments, usually all direct customers place a demand forecast in every time period.

Step 1: The error $e_{it\tau}$ of every $f_{it\tau}^{hist}$ is calculated for the past $\tau \in T^{hist}$ consecutive time periods using Equation (51). To base the customer scores on a sufficiently large dataset, T^{hist} should contain more than 30 time periods. However, in order to mitigate the risk of wrong customer scores due to changing customer behaviour over time, T^{hist} should be limited to data of maximum one year.

$$e_{it\tau} = 1 - \frac{q_{i\tau}^{hist}}{f_{it\tau}^{hist}}. \tag{51}$$

Equation (51) is based on the assumption of large industrial customers described above. Hence, $f_{it\tau}^{hist} > 0$ for all i, t and τ. For environments, in which $f_{it\tau}^{hist}$ can be 0, $e_{it\tau}$ could be formulated in accordance to the symmetric mean absolute percentage error (see Chapters 4 and 5). In this case, if $f_{it\tau}^{hist} = q_{i\tau}^{hist} = 0$, $e_{it\tau}$ has to be defined as zero as well.

Step 2: In order to separate strategic gaming behaviour from forecasting errors, which increase with the forecast horizon, the average forecast error $\overline{e}_{ih}$ for different forecast horizons h is determined with Equation (52). The horizon h is defined as the time interval between the provision of the forecast and its indicated delivery date. Customers are forecasting their demands every time period for all horizons with a maximum horizon of h^{max}.

$$\overline{e}_{ih} = \frac{1}{|T^{hist}|} \sum_{\tau \in T^{hist}} e_{i(\tau-h)\tau}. \tag{52}$$

The DDAP methodology aims at measuring the strategic component of $e_{it\tau}$. It therefore allows negative and positive values for $e_{it\tau}$ that origin from random forecasting errors to cancel each other out in Equation (52).

Step 3: To identify rationing gaming behaviour of the customers, it is tested if the values of $\overline{e}_{ih}$ are statistically significantly larger than zero. I.e. the null hypothesis $H0 : \overline{e}_{ih} \leq 0$ is tested using a suitable test statistic. With the result of this statistical test, the forecast bias b_{ih} of customer i for horizon h is determined using Equation (53). If $|T^{hist}| \geq 30$ the central limit theorem can be applied to assume that the distribution of $\overline{e}_{ih}$ can be approximated by a normal distribution. Then, the student's t-test can be used to determine the values of b_{ih}.

$$b_{ih} = \begin{cases} 0, if\ H0\ is\ accepted \\ \overline{e}_{ih}, if\ H0\ is\ denied \end{cases}. \tag{53}$$

Next, the average demand bias b_i of customer i is determined using Equation (54). In Equation (55), the forecast accuracy of a customer is defined as the complement of the average demand bias.

$$b_i = \frac{1}{h^{max}} \sum_{h=0}^{h^{max}} b_{ih} \tag{54}$$

$$acc_i = 1 - b_i \tag{55}$$

Step 4: Finally, the values of $prof_i$ and acc_i are normalised using Equation (56) and (57) and the values of PAS_i are calculated with Equation (58).

$$prof_i^{norm} = \frac{prof_i - \min_i(prof_i)}{\max_i(prof_i) - \min_i(prof_i)} \tag{56}$$

$$acc_i^{norm} = \frac{acc_i - \min_i(acc_i)}{\max_i(acc_i) - \min_i(acc_i)}; \tag{57}$$

$$PAS_i \ = (1-\alpha) \cdot prof_i^{norm} + \alpha \cdot acc_i^{norm} \tag{58}$$

6.2.2 Allocation planning model

In allocation planning, the ATP supply atp_t becoming available at the beginning of time period t is allocated to the demand forecast $f_{i\tau}$ of customer i due in period τ . The result of the process is the AATP quantity $aatp_{it\tau}$.

The allocation planning model is a modified version of the model described in Section 2.3.2. Though a heuristic, such as a greedy algorithm, could be used to find an optimal solution to the allocation planning problem, a linear programming model is chosen because of its flexibility in terms of adding additional constraints like minimum allocation quantities for customers resulting from contractual obligations of a supplier. Furthermore, the model can easily be extended with more complex constraints, e.g. for service level balancing between customers over time or allocation planning considering substitution of products. Such problems, however, cannot be solved to optimality by a heuristic anymore.

Note that the DDAP approach, which is described by Equations (59) to (62), differs significantly from model presented in Meyr (2009), which maximises the profit of the supplier. DDAP, in contrast, prioritises the demand fulfilment of preferred customers, thereby explicitly considering customer forecast accuracy in addition to profitability. Moreover, while in Meyr (2009) the supply is allocated to customer segments, the model presented here allocates it to individual customers. Finally, while the model in Meyr (2009) allows free ATP quantities, which are available for consumption by all customers, in the DDAP approach the entire available ATP supply must be allocated to the customers. Since significant scarcity of supply at all points in time is assumed, this cannot lead to infeasibilities in the model.

The objective function (59) maximises the customer-score-weighted supply allocation and penalises early and late demand fulfilment with the factors $\xi_{t\tau}^{e}$ and $\xi_{t\tau}^{l}$, respectively. It ensures that demands of the customers with high PAS_i -values are satisfied with priority. Constraints (60) ensure that the generated AATP quantities do not exceed customer demand forecasts.

Constraints (61) state that the sum of allocated supply quantities must equal the total available ATP quantities. Constraints (62) represent non-negativity constraints.

Maximise

$$z = \sum_i \sum_\tau \left[\sum_t \left(PAS_i \cdot aatp_{it\tau} \right) - \sum_{t<\tau} \left(\xi_{t\tau}^{e} \cdot aatp_{it\tau} \right) - \sum_{t>\tau} \left(\xi_{t\tau}^{l} \cdot aatp_{it\tau} \right) \right] \quad (59)$$

subject to

$$\sum_t aatp_{it\tau} \leq f_{i\tau} \qquad \forall i \in I, \tau \in T; \quad (60)$$

$$\sum_i \sum_\tau aatp_{it\tau} = atp_t \qquad \forall t \in T; \quad (61)$$

$$aatp_{it\tau} \geq 0 \qquad \forall i \in I, t \in T, \tau \in T. \quad (62)$$

6.2.3 Illustration of approach

Table 4 illustrates the potential advantages of the DDAP approach compared to the CAP approach for a simple single-period allocation planning example. CAP reserves ATP to customer segments. It thereby solely considers segment profitability and satisfies demand forecasts of segments in order of their profitability. All time indices are omitted in Table 4 because a single-period problem is investigated. Furthermore, for simplicity, the example assumes that $b_i = e_i$.

The effect of CAP and DDAP on the demand fulfilment performance is measured by using the customer service level, the profit generated from sales and the ending stock level. All customers forecast a demand f_i of 100 units. However, their order quantities q_i differ from this forecast. The sequence of order reception and the per-unit profit $prof_i$ of the customers are given in the table. A total ATP quantity of 350 units is assumed.

CAP allocates 200 units to segment 1 and 150 units to segment 2 and promises a total amount of 280 units, which leads to an ending stock level of 70 units, a total service level of 70% and a profit of 3780.

For the DDAP approach the α-level is assumed to be 0.6. The normalised customer profitability $prof_i^{norm}$ and forecast accuracy acc_i^{norm} are calculated on the basis of f_i and q_i using Equations (51) to (57) without time indices. The customer scores PAS_i follow from Equation (58). Based on these customer scores, the AATP quantities given in Table 4 are obtained when the DDAP model is run. After order promising, the DDAP approach leads to a total of 320 units of satisfied orders, an ending stock level of 30 units, a total service level of 74.3% and a profit of 3980, i.e. lower stocks, higher service level and higher profit than the CAP approach.

Such results are due to the consideration of forecast bias data on a highly granular level. Note that, for this exemplary case, DDAP approach leads to higher profits than the CAP approach even though the model does not exclusively aim at fulfilling the demands of the most profitable clients. The example further shows that DDAP approach incentivises the customers to forecast their demands truthfully, since higher values of acc_i result in higher service levels. For example, for customer 5 ($acc_5 = 0.9$), SL^{DDAP} is 100% and SL^{CAP} is 0%, while for customer 2 ($acc_2 = 0.6$), SL^{DDAP} is 0% and SL^{CAP} is 100%.

6.3 Experimental design and parametrisation

The advantages of the DDAP approach are illustrated with a numerical study using historical demand data from the semiconductor manufacturing industry. Here, supply shortage situations appear frequently due to long production cycle times, high capacity investment cost and high

Table 4: Single-product, single-period example: CAP vs. DDAP

customer i	**1**	**2**	**3**	**4**	**5**	**Total**
profitability $prof_i$ (per-unit)	15	14	13	12	11	-
f_i	100	100	100	100	100	500
q_i	70	60	90	80	100	400
ordering sequence	5	4	1	2	3	-
atp	-	-	-	-	-	350
segment k	1 (customers 1, 2)		2 (customers 3, 4, 5)			-
$prof_k$	14.5		12			-
$aatp_k^{CAP}$	200		150			350
$promise_i^{CAP}$	70	60	90	60	0	280
ending stock	-	-	-	-	-	70
SL^{CAP}	100%	100%	100%	75%	0%	70%
$profit_i^{CAP}$	1050	840	1170	720	0	3780
α	-	-	-	-	-	0.6
$prof_i^{norm}$	1.0	0.75	0.5	0.25	0.0	-
acc_i	0.7	0.6	0.9	0.8	1.0	-
acc_i^{norm}	0.25	0.0	0.75	0.5	1.0	-
PAS_i	0.43	0.24	0.81	0.62	1	-
$aatp_i^{DDAP}$	50	0	100	100	100	350
$promise_i^{DDAP}$	50	0	90	80	100	320
ending stock	-	-	-	-	-	30
SL^{DDAP}	71.4%	0%	100%	100%	100%	74.3%
$profit_i^{DDAP}$	750	0	1170	960	1100	3980

demand volatility (see e.g. Ehm et al. 2011). The customers of the industry under review display rationing gaming behaviour.

After introducing the design of experiments in Section 6.3.1, the product-individual α values to be used in the DDAP approach are determined in Section 6.3.2. The results of the numerical study are presented in Section 6.4.

6.3.1 Assumptions, data and performance measures

As depicted in Figure 14, allocation planning and order promising are run in a rolling horizon scheme. In every planning period, first, allocation planning generates AATP quantities based on ATP and customer demand forecast data. Afterwards, customer orders are realised and promised based on the allocated supply. Then, the planning horizon is rolled over, new demand forecasts and ATP quantities become available and allocation planning is performed again.

To be able to measure the capability of the DDAP approach to cope with biased demand forecasts from the customers, the following assumptions are made, which eliminate other sources of uncertainty:

1. The supply quantities atp_t are deterministic.
2. Orders will not be cancelled or rescheduled by the customers once they enter the system.

Table 5: Dataset for numerical case study

Products	P1	P2	P3	P4	P5	P6	Total
Number of customers	12	41	23	25	18	26	145
Number of orders	182	1772	723	944	727	820	5168
Average of demand biases b_i	8%	10%	5%	6%	12%	13%	9%
Share of customers with positive bias	50%	82%	48%	83%	71%	72%	67%
Average error of acc_i^{norm}($\overline{err}^{acc}$)	0.029	0.029	0.037	0.039	0.045	0.049	0.038

Furthermore, it is assumed that:

3. Orders can be fulfilled partially and with multiple shipments.
4. If a part of an order cannot be promised when it is received, that part is lost.

To measure the demand fulfilment performance, the OTSL (Equation (32)), the TSL (Equation (31)), the profit generated from sales (Equation (36)) and the average level of stock resulting from excess allocation (Equation (35)) are used. For confidentiality reasons, the real profitability of the customers is not provided in the dataset. However, information on the relation of profitabilities of the customers within the dataset is available. For the numerical study, two scenarios for the real profitabilities of customers in the dataset are assumed. In the *extreme case scenario*, the per-piece profitabilities of the most and the least profitable customers are €1 and €0, repectively; in the *realistic case scenario*, these profitabilities are € 0.1 and € 0.067, respectively.

Data from a large European semiconductor manufacturer is used. The dataset contains orders and demand forecasts for six standard products from the automotive and industrial segments of the company.

Table 5 gives an overview of the large dataset containing 78 weeks of forecast data for 145 customers and the corresponding 5168 orders, including their arrival time details. The first 52 weeks (in sample) are used to generate the customer scores PAS_i with the four-step data analysis described in Section 6.2.1. The last 26 weeks (out-of-sample) are used for the numerical study. The customers in the dataset, which the case company groups into three segments, order with an average order lead time of 3 weeks.

The positive values of the average of demand biases b_i illustrate that the customers in the dataset exhibit rationing gaming behaviour. The share of customers with a positive b_i shows that not all customers in the dataset show strategic gaming.

When calculating the forecast accuracy for the out-of-sample time period $acc_i^{norm}(out-of-sample)$, the average error $\overline{err}^{acc}$ of the historical forecast accuracy acc_i can be derived using Equation (63).

$$\overline{err}^{acc} = \frac{\sum_{i\in I}\left|acc_i^{norm} - acc_i^{norm}(out-of-sample)\right|}{|I|} \tag{63}$$

Smaller values indicate a higher predictive quality of acc_i^{norm} for the out-of-sample time period. Table 5 shows that the historical forecast accuracy values of the customers are of different predictive quality since the value of $\overline{err}^{acc}$ differs significantly between the products. However, the small values of $\overline{err}^{acc}$ show that for a typical industrial environment, the historical forecast accuracy calculated over a period of 52 weeks is usually of high predictive quality. This observation validates the approach to perform allocation planning at the individual customer level.

The customer forecast bias is calculated on a sample of 52 weeks and obtain 52 observations of the demand forecast error $e_{i(\tau-h)\tau}$ for every horizon h and every customer i. Consequently, the central limit theorem can be applied to assume that the distribution of $\overline{e}_{ih}$ can be approximated by a normal distribution. Therefore, the student's t-test is used with a significance level of 10% to determine the values of b_{ih} in Equation (53).

For the penalty costs for early and late order fulfilment, values are assigned such that early order fulfilment (i.e. temporary stock building) is preferred over late order fulfilment. Hence, $\xi^{e}_{t\tau}$ and $\xi^{l}_{t\tau}$ such that $max(\xi^{e}_{t\tau}) < min(\xi^{l}_{t\tau})$, $max(\xi^{l}_{t\tau}) < min(PAS_i)$, $\xi^{e}_{t^1\tau} < \xi^{e}_{t^2\tau}$ for $t^1 > t^2$ and $\xi^{l}_{t^1\tau} < \xi^{l}_{t^2\tau}$ for $t^1 < t^2$.

The demand fulfilment performance of DDAP and CAP is compared. A FCFS real-time order promising without preceding allocation planning serves as a benchmark. For comparability reasons, the frequency of the CAP approach is set to the frequency of the DDAP approach, i.e. one week.

Note that the DDAP approach is developed for the single product case. Hence, when applying it to multiple products like in this numerical study, the values of PAS_i and α can be determined separately for all products.

The numerical study is implemented in Java. IBM ILOG CPLEX V12.6.0 is used to solve the linear programming models for allocation planning and order promising. The study was performed on a personal computer with an Intel Xeon E7-4860 v2 processor with 2.6 GHz and 32GB RAM on a 64-bit Microsoft Windows 7 installation.

6.3.2 Trade-off between profitability and forecast accuracy

To be able to investigate the effects of considering customer forecast accuracy in allocation planning, first, the α values to be used for each product dataset in the numerical study need to be determined. For this calibration of α, first, the PAS_i values for each dataset are determined using the 52 weeks in-sample data and the DDAP method is run on the in-sample data varying the level of α between 0 and 1 in five equidistant steps. The analysis is done at a supply shortage level of 20%, which is defined as the level to which the total customer demand exceeds the total available supply. The determined α values are denoted by α^*.

In the following, the determination of α^* for P4 is demonstrated in detail. For all other products, the results look similar and the determination of α^* is done in the same way. The determined α^* values for all six products are provided at the end of the section.

Figure 15 and Figure 16 illustrate the influence of the α level on the service level, profit and average stock level resulting from excess allocations for P4. Figure 15 shows the overall service levels as well as the performance for the first quartiles of customers with the lowest and highest forecast biases. The results are shown relative to the DDAP performance at an α level of 0.

Intuitively, the TSL and the OTSL for customers with low forecast bias increase with increase in α, because demand fulfilment prioritises these customers more as α increases. Analogously, the service levels decrease for the customers with high forecast biases. However, the increase in the service levels for low bias customers is significantly higher. It is because the majority of the customers with low biases have a low profitability and order relatively small volumes.

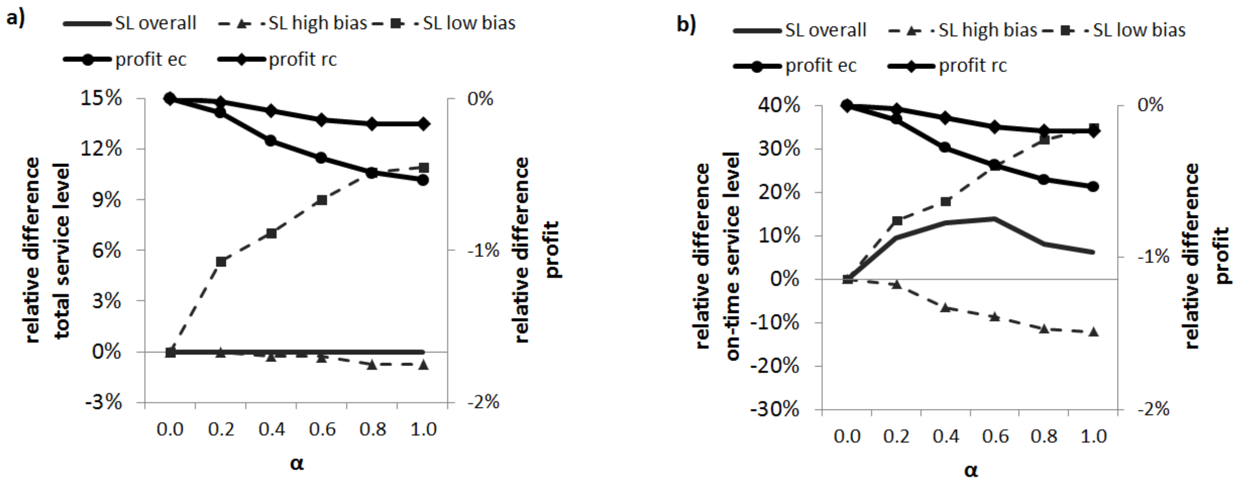

Figure 15: a) total service level and b) on-time service level and profits in dependence of the level of α

Therefore, their demands are not satisfied at all when the α level is small. When α increases and their demands are satisfied, their service level increases significantly. On the other hand, the customers with a high forecast bias mainly order high volumes and have high profitability. Therefore, even though a smaller portion of their demand is satisfied at higher levels of α, their orders are still fulfilled partially leading to a smaller decrease in the service level.

Figure 15 further shows that the overall TSL does not depend on the level of α. This is because the fulfilment of orders after their requested delivery date is allowed, AATP quantities are nested and problems in which ATP is scarce over the entire planning horizon are studied. Therefore, the entire ATP supply is always consumed, which leads to an α-level-independent overall TSL.

The overall (average) OTSL first grows for $\alpha < 0.6$, then reaches a maximum and decreases monotonically for $\alpha > 0.6$. This has two reasons. First, prioritising the customers that forecast their demand truthfully reduces the risk of excess allocation. Hence, the risk of late order promising while at the same time generating temporary stocks is reduced. As a result, ATP supply can be consumed more efficiently, leading to a higher OTSL performance. Second, the customers with low demand biases tend to place their orders later than others. When placing large emphasis on forecast accuracy (high α), the supply allocation for these customers increases. Even though their forecasts are more accurate, they still include a demand bias. However, due to the short lead time, this bias cannot be compensated for by making the excess allocation available for other orders. Therefore, even though the OTSL for the customers with low forecast bias increases with increase in α, the overall OTSL deteriorates for α levels above 0.6.

For the same reason, the average stock levels resulting from excess allocation show a minimum at an α level of 0.6 (see Figure 16). The consideration of forecast bias data in demand fulfilment, thus, additionally reduces holding cost for inventory resulting from excess allocation.

Going forward, the α level that maximises the OTSL and minimises the average stock level is called *SL-optimal* α. As the above discussion shows, the location of the SL-optimal α partly depends on the distribution of the order lead times in the customer set, i.e. the time between order placement and the requested delivery date.

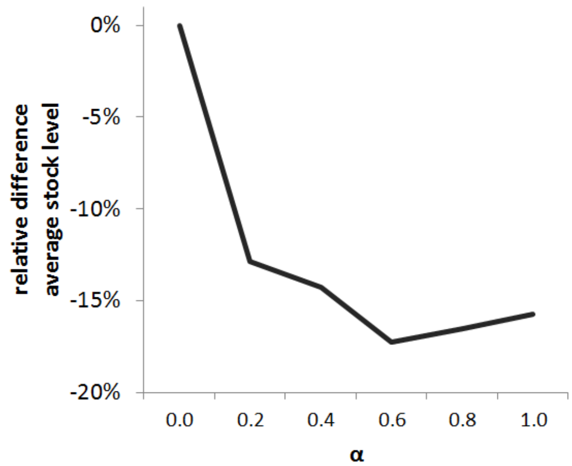

Figure 16: Average stock resulting from excess allocation

Since the profits realised are decreasing with increase in α, Figure 15 shows a trade-off between maximising profits and maximising the overall service levels and minimising the average stock levels. The trade-off exists because the highly profitable customers also show the highest demand bias. In practice, it can be explained by a higher market power that the highly profitable customers exercise over their suppliers. Being aware of their strong position, the customers strategically game their suppliers' allocation planning.

For P4, the average decrease in profit at the SL-optimal α level is only 0.14% and 0.39% for the realistic case scenario and the extreme case scenario, respectively. Both values are very small, especially for the realistic case scenario, which shows differences in customer profitability that are comparable to the industry. For P4, α^* is therefore set to the SL-optimal α, i.e. 0.6.

For the other five products in the dataset, the same analysis is conducted. For all products the results look similar. Only the SL-optimal value of α differs. As a consequence, the following levels of α^* are obtained for the products P1 to P6: 0.6, 0.2, 0.4, 0.6, 0.8 and 0.4.

Obviously, the determined value of α^* represents the trade-off made between maximising profits and maximising service levels. Note that for other cases, especially when the choice of α affects the total profits more than in the case investigated here, the choice of α^* can differ from the SL-optimal α.

6.4 Numerical results

This section presents the results of the numerical study defined in Section 6.3. The benefits of considering forecast bias data in allocation planning and order promising are analysed in Section 6.4.1. Section 6.4.2 contains an analysis that demonstrates that the DDAP approach incentivises the customers to provide truthful forecasts for the given dataset. Then, the performance of DDAP and CAP is compared in Section 6.4.3. In Section 6.4.4, the analysis is concluded with investigating the effects of moving the demand fulfilment level from customer segment to individual customer and considering demand bias data separately.

6.4.1 Benefits of considering forecast bias data

To investigate the benefits of considering forecast bias data in demand fulfilment, the DDAP approach is used on the out-of-sample data and the level of supply shortage is varied from 10%

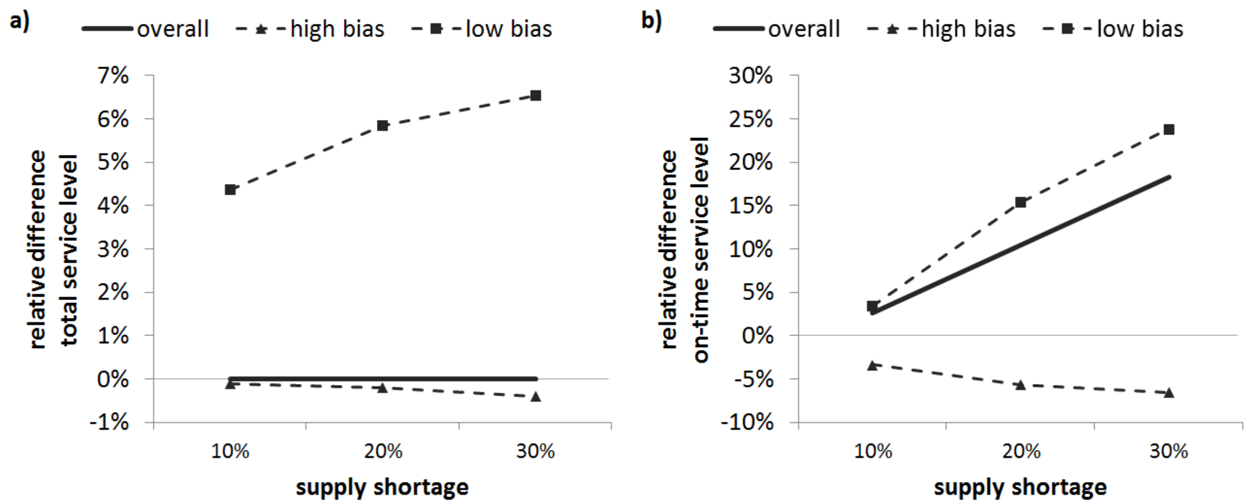

Figure 17: a) total service level and b) on-time service level in dependence of the level of supply shortage

to 30% in two equidistant steps using, first, the α^* values determined in Section 6.3.2 and, second, an α level of 0.

Figure 17 shows the service levels of the DDAP approach at different levels of supply shortages and their respective α^*. The findings are displayed relative to the performance of the DDAP approach at an α level of 0, i.e. only considering customer profitability. The graphs show that the positive effect of considering forecast accuracy data on the overall OTSL increases with the level of supply shortage. The reason for the amplification of the given effects lies in the scarcity of supply itself.

As shown above, supply can be allocated most efficiently at α^*. Moreover, the influence of allocating supply more efficiently among customers on the service levels grows with the level of supply shortage. While at low supply shortage levels relatively many orders of the customers with lower scores can be fulfilled even though excess allocation exists, lesser orders can be fulfilled as supply keeps becoming scarce. Hence, when allocating supply more efficiently, the additional number of satisfied orders from the customers with lower scores compared to demand fulfilment without consideration of demand bias data grows with the growing level of supply shortage.

6.4.2 Impact of truthful forecasting for highly profitable customers

It was shown above that the DDAP approach increases both the TSL and OTSL for the customers with high forecast accuracy. On the other hand, both the service levels decrease for the customers with low forecast accuracy. In this section it is shown that using this effect, the DDAP approach incentivises all the customers to increase their forecast accuracy.

In order to realise this incentive, suppliers need to communicate the effects of DDAP on the individual service levels of customers depending on their forecast accuracy. It is obvious that customers with relatively low profitability have an incentive to increase their customer score PAS_i by sustainably improving their forecast accuracy to increase their priority in the allocation planning process of the supplier and consequently increase the service levels with which they are served.

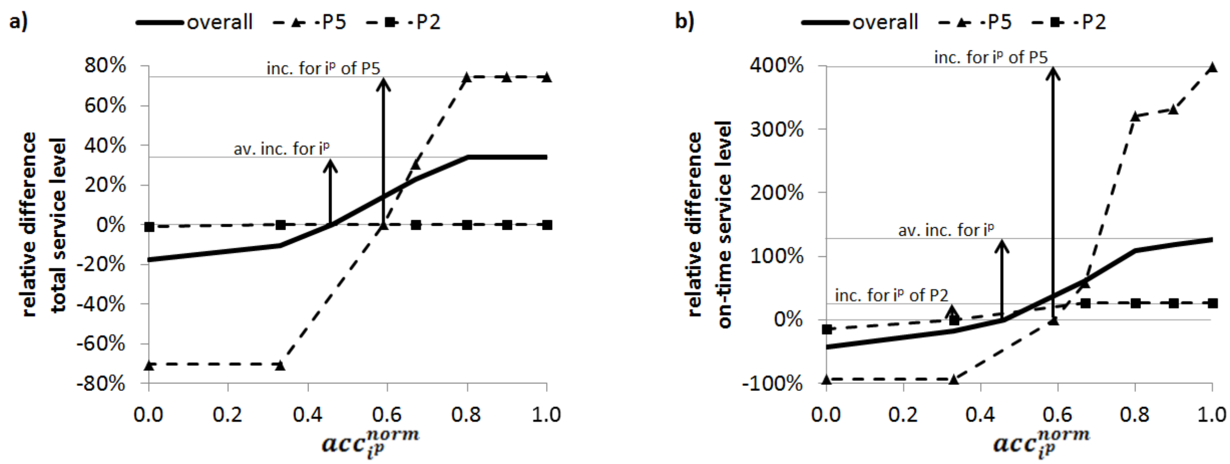

Figure 18: Effect of forecast accuracy on a) TSL and b) OTSL of the most profitable customer

However, the numerical results show that in supply shortage situations the TSL and OTSL for all the customers are smaller than 100%. In the following it is investigated whether such an incentive also exists for customers with relatively high profitability. For this, the forecast bias and the forecast accuracy acc_i of the most profitable customer i^p for all six product datasets is varied between 0 and 1, their forecasts are adapted in the out-of-sample data and the effect of this variation is analysed on the TSL and OTSL of these customers.

Figure 18 shows the effects of the variation of acc_{i^p} on the TSL and OTSL of i^p as an average across all six product datasets (overall) and for the two products with the largest (P5) and the smallest (P2) α^* (see Figure 16). P5 and P2 are chosen, because they show the largest and smallest effects. All other products display the same pattern, but the effect of acc_{i^p} on the service levels is differently strong.

The graphs show the relative difference between the service levels at certain forecast accuracy and the service level at the real forecast accuracy of i^p in the respective datasets. The vertical arrows indicate the potential improvement of TSL and OTSL for i^p if they increase their normalised forecast accuracy to 100%. This potential represents the incentive for the customers to improve their forecast accuracy. Note that a normalised forecast accuracy of 100% does not mean a bias-free demand forecast, rather the lowest bias within the customer dataset.

On average (solid line), the potential relative improvements of the TSL and OTSL for the most profitable customers are about 30% and 120%, respectively. Hence, the customers are strongly incentivised to increase their forecast accuracy. Due to the high α^* of 0.8, these incentives are even higher for the most profitable customer of P5. The potential relative improvements of TSL and OTSL are about 75% and 400%. For the most profitable customer ordering P2 for which α^* is low (0.2), the potential service level improvements are comparably low. However, the customer can still improve their OTSL by about 35% provided they increase their normalised forecast accuracy to 100%.

These numerical findings show that it is not a dominant strategy for the most profitable customers in the dataset to inflate their demand forecasts. All investigated customers are incentivised to increase their forecast accuracy and, hence, forecast demands truthfully, if the supplier employs the DDAP approach.

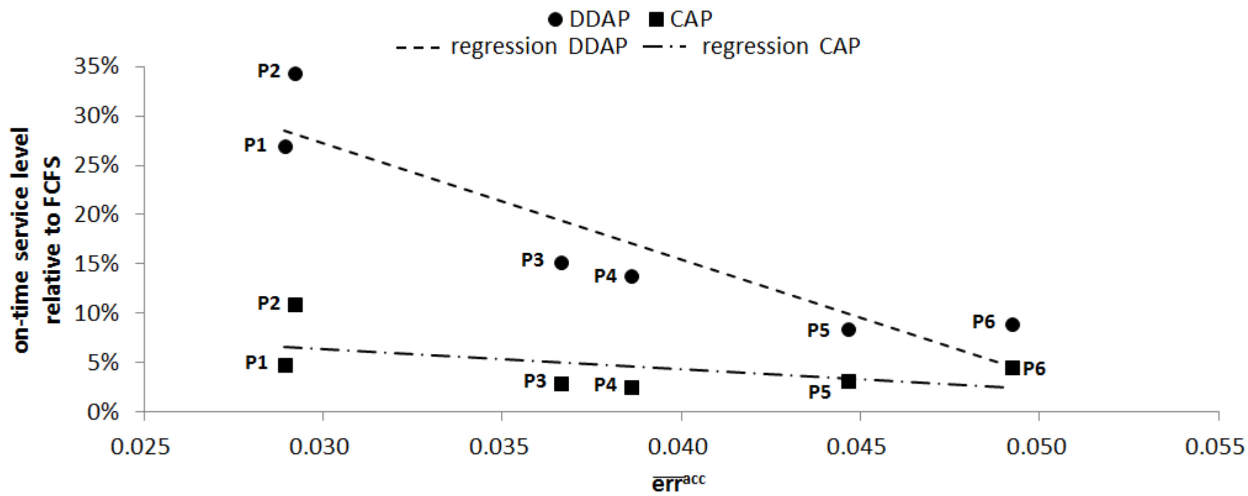

Figure 19: Service levels of DDAP and CAP in dependence of the predictive quality of data

6.4.3 Comparison of demand fulfilment approaches

Figure 19 depicts the OTSLs of DDAP and CAP relative to the FCFS approach in dependence of the average error of the historical forecast accuracy values $\overline{err}^{acc}$, i.e. the predictive quality of the historical forecast accuracy values. Like the approach in Meyr (2009), CAP allocates supply to customer segments solely based on customer profitability data. The models used for allocation planning and order promising in the CAP approach are presented in Sections 2.3.2 and 2.3.3. The case company groups their customers into three different customer segments. These are used in the CAP approach.

The graph shows that the DDAP approach outperforms the CAP approach for all the studied product data sets, i.e. predictive qualities of the historical forecast accuracy values, because of its more efficient supply allocation; thus, exploiting the availability of highly granular data. However, the advantage decreases with increasing $\overline{err}^{acc}$. This is intuitive since high levels of predictive quality indicate stable forecasting behaviour of the customers. Consequently, the SL-optimal α values, which are chosen as α^*, do not change for the out-of-sample data, on which the analysis is based. On the other hand, in a situation where $\overline{err}^{acc}$ is high, the customers' forecasting behaviour cannot be predicted accurately and the potentials of highly granular data cannot be exploited as efficiently by the DDAP approach.

6.4.4 Effects of demand fulfilment on customer service level and consideration of demand bias

The DDAP approach differs in two respects from the CAP approach. First, allocation planning and order promising are done on individual customer level. Second, the demand bias data is taken into account. In this section, the effects of both these aspects are investigated separately.

To be able to measure the effect of considering forecast bias data when fulfilling demands at the customer segment level, the DDAP model is adapted to do allocation planning and order promising on customer segment level. For this, first, the average of the in-sample forecast bias b_k and the profitability $prof_k$ of the customer segment k are calculated by taking the average of the b_i values and the $prof_i$ values for all the customers i belonging to the segment k. Here, again, the customer segments of the case company are used. Afterwards, the normalised

Table 6: Effects of demand fulfilment granularity and consideration of demand bias on the on-time service level

	product	customer segments OTSL	customer segments profit	individual customers OTSL	individual customers profit
CAP	P1	-	-	+0.0%	−0.01%
($\alpha = 0$)	P2	-	-	+0.9%	0.01%
	P3	-	-	+1.0%	−0.01%
	P4	-	-	+0.1%	0.00%
	P5	-	-	+0.3%	0.00%
	P6	-	-	+1.1%	0.01%
DDAP	P1	3.3%	−0.03%	8.9%	−0.13%
(α = SL-optimal)	P2	2.4%	−0.02%	9.3%	−0.12%
	P3	4.0%	−0.07%	5.0%	−0.09%
	P4	3.1%	−0.02%	5.1%	−0.06%
	P5	2.2%	−0.01%	2.4%	−0.02%
	P6	3.2%	−0.03%	1.5%	−0.02%

forecast accuracy acc_k^{norm} and profitability $prof_k^{norm}$ of the segments as well as the segment scores PAS_k^{seg} are generated using Equations (56) to (58) for customer segments. Finally, the so generated PAS_k^{seg} are used as $score_k^{seg}$ in the allocation planning model described in Section 2.3.2.

Table 6 presents the effects of fulfilling demands at customer segment or individual customer levels as well as using profitability of customers for customer scores alone or additionally considering demand bias data. When reading the table from left to right, the effect of changing the demand fulfilment granularity from customer segment to individual customer is shown. Reading the table from top to bottom shows the effect of considering forecast bias information in customer scores. All numbers are relative to the performance of the CAP approach.

Changing the demand fulfilment granularity from customer segment to individual customer increases the OTSL between 0.0 and 1.1 and the profit (real case scenario) between −0.01 and 0.01 percentage points. Therefore, changing the demand fulfilment granularity alone only has a weak effect on demand fulfilment performance. This is because, first, pooling effects within customer segments and, thus, flexibility in order promising is lost. Second, the effect on the OTSL is positive due to the fact that most customers in the dataset display a tendency of inflating their demand forecasts, and the highly profitable customers place their orders earlier than others. Hence, excessive AATP quantities for the highly profitable customers can be used to fulfil orders of the less profitable customers, which are received later.

Considering demand bias data when fulfilling demands at the customer segment level increases the OTSL moderately between 2.2 and 4.0 percentage points. Profits are decreasing slightly between −0.07 and −0.01 percentage points. The reason why the benefits in terms of OTSL are not more significant is that customers within these profit segments are very heterogeneous in terms of their individual forecast accuracy. As a consequence, customers with high demand biases are grouped with customers forecasting their demands truthfully. Raising the α level leads to a prioritisation of customer segments with low average demand bias. Nevertheless, the demand biases of individual customers within prioritised segments can still be high. The overall profit decreases since, on average, customers with higher forecast accuracy tend to have lower profitability.

Finally, combining both these aspects results in the highest service level for all products except P6. The most distinct effects are achieved for the products P1 and P2, for which the OTSLs of both the DDAP approach on customer segment level and the CAP approach on individual customer level are significantly outperformed. Analogously, a decrease in the overall profits is most pronounced for these products. For the products P3 to P5, the increase in OTSL compared to the CAP approach at the individual customer level is moderate while the advantages compared to the DDAP approach at the customer segment level are low. The overall profits change analogously.

The rationale behind the different magnitudes of leveraging effect is the difference between the predictive qualities of forecast accuracy values. While these values are of high predictive quality for the product datasets P1 and P2, they are of moderate quality for P3 to P5, and give only little indication of the forecast accuracy of customers out-of-sample in the dataset P6. The higher the predictive quality of the forecast accuracy values, the less temporary are the stocks resulting from excess allocation that are built by the DDAP approach. In dataset P6, $\overline{err}^{acc}$ is exceptionally high so that the usage of forecast accuracy data at the individual customer level leads to a negative effect on the OTSL compared to the DDAP approach at the customer segment level.

We, therefore, draw the following conclusions. First, the use of the DDAP approach is only meaningful when additional data at the individual customer level is considered. Second, when looking at OTSL, allocation planning at the individual customer level, considering demand bias data robustly outperforms the conventional allocation planning approaches, allocating supply to customer segments purely based on profitability data. In exceptional cases, when the predictive quality of the historical forecast accuracy is very low, it is, however, more beneficial to fulfil demands at the customer segment level. Third, the benefits of the DDAP approach increase with the predictive quality of the historical data on customer forecast accuracy.

6.5 Summary and conclusion

The recent advances in big data analysis tools represent an opportunity for companies to further improve their demand fulfilment processes. In particular, the exploitation of data on the ordering behaviour of customers, e.g. their forecast biases, can enable companies to improve the accuracy and robustness of their order promises and increase the performance of their demand fulfilment systems. Here, a first step towards integrating such newly available data into allocation planning and order promising has been taken.

Research has shown that customers systematically inflate their demand forecasts in supply shortage situations to make suppliers increase the supply quantities reserved for them and finally satisfy their total actual demand. This behaviour called the rationing gaming significantly impairs the ability of the supplier to efficiently allocate current and future supply to customers.

An allocation planning methodology called the data driven allocation planning (DDAP) is proposed, which considers the data on individual customers and products by allocating supply on a highly granular level, taking systematic biases in demand forecast into account. The DDAP approach supports an efficient supply allocation, leading to a significant increase in customer service levels compared to conventional allocation approaches. In addition, the approach reduces stocks that are created by excess allocation to customers showing low forecast accuracy. By increasing service levels for customers with high forecast accuracy, the approach incentivises customers to provide truthful demand forecasts and, thus, counteracts rationing gaming.

The DDAP approach is tested in a numerical study using data from a large European semiconductor manufacturer. The analysis proves that allocating supply to individual customers is only valuable when additional data on this granularity level is available and taken into account. The results demonstrate the capability of the proposed approach to significantly increase allocation efficiency and incentivise customers for truthful forecasting for the investigated case. The improvements become more distinct with growing scarcity of supply and increased predictive quality of the available data.

This work uses an extensive dataset from the semiconductor industry. Considering the special characteristics of this industry; i.e. high volatility of demand, long production lead time, capital intensive manufacturing and particularly short product life cycles; it is reasonable to believe that the DDAP approach is also beneficial in other industries, which are typically less dynamic. Nonetheless, to prove its benefits in general also for other industries, this work could be complemented by testing the applicability of the approach in different environments. In this context, especially the optimal weight factor for balancing forecast bias and profitability of customers and a general analytical proof of the incentive for customers to forecast their demands truthfully would warrant further investigation.

The findings of the study presented here indicate that an allocation approach additionally accounting for data on customer order lead times could lead to even better results. Thus, for further research purposes, it would be interesting to integrate order lead time data on the individual customer level into allocation planning approaches. Additionally, approaches to increase the performance of DDAP in environments with varying predictive quality of historical data have to be developed. Allocation planning methods switching from an individual customer level to a customer segment level when the volatility of demand biases reaches a certain threshold should be investigated. Moreover, considering substitute products in the allocation planning decision would be interesting; especially, taking into account data on the individual willingness of customers to substitute products has so far not been addressed. Finally, the examination of the performance of DDAP under supply uncertainty or demand uncertainty after order arrival indicates possibility of a further extension. Hereby, order rescheduling and cancellation rules, which industrial suppliers and customers agree upon in supply contracts, have to be taken into consideration. Obviously, several of these further research directions can be realized by exploiting additional data.

7 Managing the contract portfolio to increase demand fulfilment performance

When demands exceed capacities, suppliers allocate available supply to customers based on customer importance and advance demand information. The accuracy of advance demand information interacts with the length of customer order lead times and influences the overall customer service levels.

In this chapter, industrial contract portfolios with customer-specific terms are analysed in order to derive insights aiding suppliers in their contract portfolio management and in their design of demand fulfilment processes. For this purpose, a framework is developed for an analysis of contract portfolios capturing the dynamics of industrial planning processes. The framework is applied to portfolios from the semiconductor sector.

A numerical analysis shows that, in order to improve service levels, an industrial demand fulfilment process must take all contract terms, including order lead times and historical forecasting and ordering behaviour, of all customers into account.

In general, suppliers prefer long order lead times. However, the analysis shows that demand fulfillment performance is not primarily determined by the absolute length of the order lead times but by the presence of a negative correlation with the accuracy of advance demand information in the entire contract portfolio. Consequently, suppliers must consider the portfolio of all customers and negotiate relatively long order lead times for customers showing relatively low accuracy of advance demand information.

In the following, Section 7.1 describes the use of ADI in industrial demand fulfilment and illustrates the contributions of the work presented in this chapter. Section 7.2 explains the allocation planning approach in detail. Section 7.3 introduces a case from the semiconductor industry, describes the experimental design of the numerical study and presents its results. Section 7.4 gives managerial implications and concludes the chapter.

7.1 Advance demand information in demand fulfilment

In environments in which production cycle times exceed customer order lead times, i.e. the time between order placement and the requested delivery date, the move towards mass customisation has led to a high risk of too long demand fulfilment lead times and excessive stocks (see e.g. Giard and Mendy 2008). Companies therefore share so-called ADI, which are forecasts of future demands that customers provide to their immediate upstream suppliers (see e.g. Hariharan and Zipkin 1995). The literature dealing with ADI usually assumes enough supplier capacity to fulfil the total forecasted demand in the forecast horizon. In this case, suppliers feed the ADI into their demand planning process, where it is consolidated with internal forecasts (see left side of Figure 20) and forwarded to the supply network planning process, which plans supply chain activities accordingly (see e.g. Karrer et al. 2012 or Altendorfer and Minner 2014). The resulting supply information is then passed on to the demand fulfilment process, where it is used to promise delivery dates for incoming customer orders in real-time (see Chapter 5). For such situations, previous research has shown that ADI increases the supply chain performance by reducing the planning uncertainty of the supplier and increasing on-time service levels at lower inventory levels (see e.g. Thonemann 2002).

The time horizon over which ADI is shared is usually not longer than a few months. Capacity planning decisions, however, are made on longer time horizons. Hence, the exchange of ADI does not mitigate the risk of investing in the wrong amount of capacities. Consequently, demand often exceeds capacities. Then, suppliers usually do not use ADI to improve their

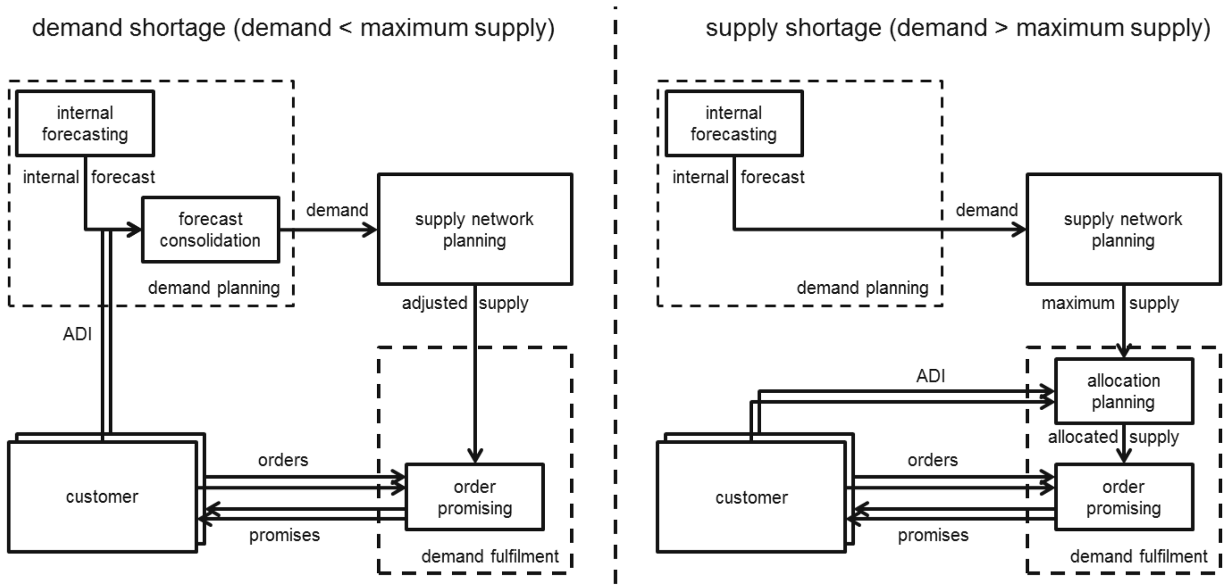

Figure 20: ADI interface to the planning process for a product group (for demand shortage and supply shortage).

internal forecasts, because any produced volumes can be sold (see right side of Figure 20). Instead, supply is planned so that profits are maximised and strategic requirements are met. The supply is allocated to customers based on the provided ADI while considering customer importance. Afterwards, the allocated supply is forwarded to the order promising process.

Anticipating such allocation mechanisms, customers strategically inflate their ADI in order to influence the allocation. This behaviour, called rationing gaming, results in the risk of low on-time service levels and high inventories due to excess allocations (see e.g. Lee et al. 2004). As the ADI inflation, also called bias, is usually not constant over time, suppliers cannot discount the ADI by a fixed value in order to improve its accuracy (Chapter 6). Consequently, suppliers aim to reduce planning uncertainty by substituting ADI through orders as early as possible, i.e. by increasing the order lead time of customers. However, because of their own demand uncertainty, customers are usually not willing to place orders long in advance.

This conflict of interests is dealt with in contract negotiations between customers and suppliers. Supply chain partners typically agree on two separate contracts. One contract sets the terms and conditions for the ADI exchange including horizon and maximum volatility. The other contract regulates the ordering process and sets i.a. the minimum order lead time and the maximum deviation between order and ADI. Hence, the contractual terms agreed upon set the boundary conditions for the demand fulfilment process of the supplier.

Since the negotiated contracts are different for different customers and the demand fulfilment process of the supplier must decide whether an incoming order from a customer with a long order lead time should be accepted or if the order should be declined to reserve the supply for other customers based on their (possibly biased) ADI, the contractual terms for the minimum order lead times and the accuracy of the ADI interact not only for one customer but also across different customers. Therefore, when negotiating new dyadic contracts, the order promising and allocation rules and the ADI and order lead time terms in the contract portfolio must be considered simultaneously. This highly relevant problem has so far not been addressed in academic publications.

The work presented in this section has the following contributions:

- Industrial contract portfolios with customer-individual terms for order lead times and ADI are analysed in order to derive insights for portfolio management.
- A framework is developed that extends the demand fulfilment methodology presented in Chapter 6 to investigate contract portfolios in the dynamic context of industrial planning processes.
- It is demonstrated how to apply and parametrise the framework to contract portfolios from the semiconductor sector and show that customers should not be grouped into segments but receive allocations individually.
- The numerical analysis shows that the consideration of order lead times results in significant improvements in service levels and performance robustness.
- The numerical results show that those portfolios, in which order lead times and the accuracy of ADI are negatively correlated, perform substantially better. Such portfolios are even superior to portfolios in which all customers have long order lead times. They allow a more efficient reallocation of excess allocations.

The main managerial implications are:

- Demand fulfilment approaches should not cluster customers in segments, but allocate ATP to individual customers, taking the individual lead times and ADI accuracy into account. Thereby it is important to know the predictive quality of the data used.
- Order lead times of a portfolio are ideally distributed such that relatively long order lead times are negotiated with customers providing ADI with relatively low accuracy.

7.2 A flexible demand fulfilment framework for evaluation of ADI and order lead time contracts

In this section, a new demand fulfilment framework is developed, which is used as a testbed for industrial contract portfolios. The framework considers the contractual parameters set for each individual customer as well as the resulting customer forecasting and ordering behaviour. It extends the methodology presented in Chapter 6 by (1) considering order lead times, (2) using a dynamic updating process for customer scoring to reflect changing customer ordering behaviour, and (3) adding flexibility to the supply allocation and order promising processes. Its processes are executed in a short-term rolling horizon fashion (e.g. every week). Figure 14 provides an overview.

Based on profitability, order lead time and ADI data, the customer score determination process determines the importance, i.e. the customer score $score_i^{cust}$, of every individual customer $i \in I$ in demand fulfilment (see Section 7.2.1).

The customer segmentation process (Section 7.2.2) segments customers into a given number of segments $|K|$ ($k \in K$) that, together with their respective segment score $score_k^{seg}$, are used in the allocation planning and order promising processes.

In allocation planning (Section 7.2.3), the ATP supply atp_t becoming available at the beginning of planning time period t is allocated to the customer segments k. The allocation is based on ADI demands $f_{i\tau}$ of the customers i being forecasted for period τ. The resulting AATP supply $aatp_{kt\tau}$ is used by a real-time order promising process to generate order promises $p_{it'\tau'}$ for incoming orders with order quantity $q_{i\tau'}$. Here, the delivery period and the time period of the requested delivery date are denoted by the indexes t' and τ'.

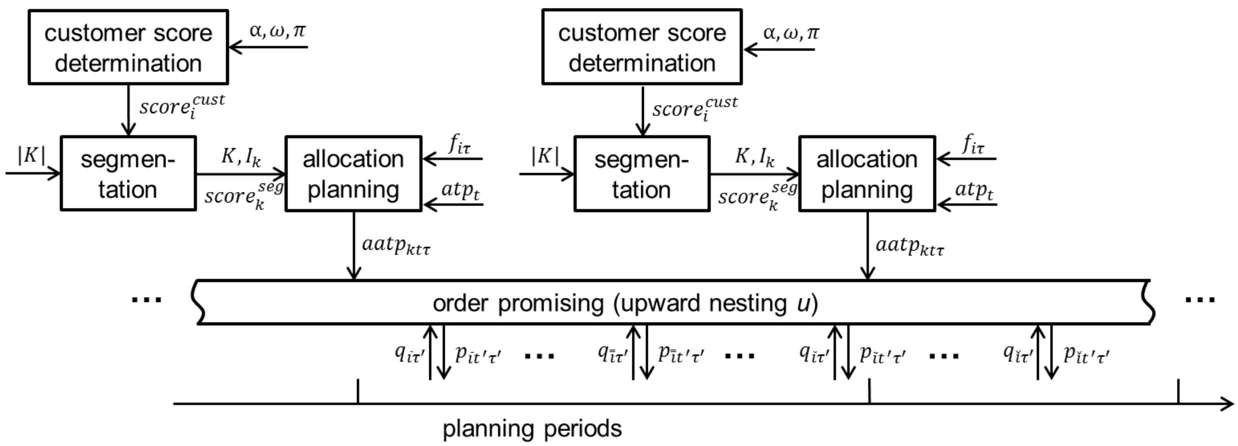

Figure 21: Rolling horizon scheme for customer ordering behaviour driven allocation planning

The order promising step allows for upward nesting in order to increase order promising flexibility (see Section 7.2.4). Orders can consume AATP quantities reserved for customers with lower priority. Additionally, allocations of u customers with higher priority can be consumed.

7.2.1 Customer score determination

The customer score determination approach uses the normalised profitability $prof_i^{norm}$, ADI accuracy acc_i^{norm}, and average order lead time olt_i^{norm} of the customers to determine their priority in demand fulfilment.

The parameters $prof_i^{norm}$ and acc_i^{norm} are derived as presented in Chapter 6. To determine olt_i^{norm}, Equations (64) and (65) are used. The time period of the due date and the time period of the placement of order o are represented by t_o and t_o^{placed}. The set O_i contains all orders o of customer i in the past $t \in T^{hist}$ time periods. Note that the boundary conditions for olt_i are set by the contracts the supplier closes with their customers.

$$olt_i = \frac{\sum_{o \in O_i} \left(t_o - t_o^{placed} \right)}{|O_i|}. \tag{64}$$

$$olt_i^{norm} = \frac{olt_i - \min_i(olt_i)}{\max_i(olt_i) - \min_i(olt_i)} \tag{65}$$

The customer scores $score_i^{cust}$ are determined by Equations (66). The factors α, ω, and π determine the weight of the historical ADI accuracy, order lead time, and profitability of a customer in $score_i^{cust}$.

$$score_i^{cust} = \alpha \cdot acc_i^{norm} + \omega \cdot olt_i^{norm} + \pi \cdot prof_i^{norm} \tag{66}$$

7.2.2 Customer segmentation

For customer segmentation, the model presented in Section 2.3.1 is used. The minimum segment size s_{min} is set to the value of $\left\lfloor \frac{2-|I|-|K|}{2(1-|K|)} \right\rfloor$, which sets a rather large minimum segment size, but still leaves flexibility to the customer segmentation model to optimise the overall maximum segment width $\overline{w}$ (see Appendix A). The value is chosen because industrial companies usually aim at levelling out the sizes of the customer segments in demand fulfilment. Note that $|K| \geq 3$ is assumed, for which $s_{min} = \left\lfloor \frac{2-|I|-|K|}{2(1-|K|)} \right\rfloor$ always leaves at least $(s_{min} - 1)$ customers unassigned (see Appendix A).

After customer segmentation, the set of customer segments K, the sets I_k containing the customers belonging to the segment k and the segment scores $score_k^{seg}$ are provided to the allocation planning process. The segment scores are calculated with Equation (67), in which v_{ik}^* are the optimal values of the binary decision variables v_{ik}, which assign assign customer i to segment k.

$$score_k^{seg} = \frac{\sum_{i \in I}(v_{ik}^* \cdot score_i^{cust})}{\sum_{i \in I} v_{ik}^*} \tag{67}$$

Note that, if $|K| = |I|$, each customer segment contains only one customer and the subsequent allocation planning and order promising processes will be performed on individual customer level.

7.2.3 Allocation planning model

The allocation planning process of the framework is a modification of the approach in Chapter 6. Its mathematical representation is presented in Section 2.3.2. The model uses the segment scores $score_k^{seg}$ that consider customer profitability and ordering behaviour data.

7.2.4 Order promising

For order promising, the model presented in Section 2.3.3 is used. The model allows nesting of customer segments. The segments, from which customer i^* is allowed to consume allocated supply, are represented in the set K_{i^*}. This set contains all segments for which Inequality (68) holds. The function RNK ranks the customer segments in descending order of their segment scores into the interval $[1; |K|]$. The customer segment of the ordering customer is represented by k^*. The upward nesting level u determines the number of customer segments with higher $score_k^{seg}$, whose AATP quantities k^* can consume.

$$\left[RNK\left(score_k^{seg}\right) + u\right] \geq RNK\left(score_{k^*}^{seg}\right) \tag{68}$$

In the numerical study presented below, fulfilment of orders before their due date is not allowed. Therefore, the promises $p_{it'\tau'}$ are calculated with Equations (29) and (30) in Section 2.3.3.

7.3 Performance analysis of contract portfolios from the semiconductor industry

In this section, different customer contract portfolios are analysed using data from the semiconductor industry. The contract portfolios are designed so that the resulting customer ordering behaviour shows different correlations of the length of the order lead time and the accuracy of ADI in the customer set.

After introducing the design of experiments in Section 7.3.1, the demand fulfilment framework presented in Section 7.2 is parametrised in Section 7.3.2. In Section 7.3.3, a numerical analysis on the impact of exact parametrization of the approach is performed, its

Table 7: Contract portfolios for numerical case study

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	Total
Customers	16	12	13	13	16	14	14	16	15	14	143
Orders	105	98	131	136	197	185	153	210	214	215	1644
Average $\overline{b}_i$	15%	21%	18%	7%	9%	12%	32%	13%	9%	1%	14%
Average olt_i	2.20	3.38	3.07	1.87	1.47	3.44	2.98	2.82	3.15	2.98	2.73
Average $\overline{err}^{acc}$ (in-sample)	0.03	0.19	0.27	0.31	0.41	0.38	0.40	0.44	0.51	0.65	0.36
Average $\overline{err}^{acc}$ (out-of-sample)	0.04	0.21	0.29	0.30	0.38	0.39	0.40	0.46	0.48	0.59	0.35
$r(olt_i, acc_i)$	0.28	-0.42	-0.18	0.43	0.05	-0.39	0.11	-0.03	0.32	0.23	0.04

performance is compared to other demand fulfilment methodologies and the influence of different contract portfolio designs on the demand fulfilment performance is analysed.

7.3.1 Assumptions, performance measures, and contract portfolios

The approach presented in Section 7.3.2 is implemented as shown in Figure 14. The following two assumptions are made to eliminate all sources of uncertainty other than ADI bias:

- ATP quantities are deterministic.
- Orders will not be cancelled or rescheduled by customers once they enter the system.

Furthermore, two assumptions on the demand fulfilment processes are made:

- Orders can be fulfilled partially and with multiple shipments.
- If a part of an order cannot be promised when it is received, this part is lost.

The effect of the design of contract portfolios on the demand fulfilment performance is measured in order to derive insights for portfolio management. The chosen indicators are the OTSL (Equation (32)) and the total profit (Equation (36)).

The dataset contains 78 weeks of ADI for 143 customers and the corresponding 1644 orders for ten products (Table 7). The design of the contract portfolio differs for all products. As a result, the length of the order lead time and the accuracy of ADI (acc_i) show different correlations $r(olt_i, acc_i)$ for all products. Furthermore, the demand data for all ten products differs in its predictive quality, i.e. the degree to which the future customer ADI accuracy can be predicted from the past. To measure the predictive quality of data, the average error $\overline{err}^{acc}$ of the ADI accuracy (Equation (69)) is used. It calculates the average of the actual error in every period t. Here, $acc_i^{norm}(t)$ is the (normalised) actual accuracy of customer i in period t and acc_i^{norm} is the (normalised) accuracy of customer i calculated with historical demand data. To calculate acc_i^{norm}, a history of 30 periods before t is used. Smaller values of $\overline{err}^{acc}$ indicate a higher predictive quality of data. In the remainder, the term contract portfolio is used as a synonym for a product of the dataset.

$$\overline{err}^{acc} = \frac{\sum_{i \in I} \sum_{t=53}^{78} \left| acc_i^{norm} - acc_i^{norm}(t) \right|}{|I| \cdot |T|} \tag{69}$$

In the numerical study presented below, the weeks 31 to 52 of the data set are used to parametrise the approach (in-sample). The experiments are then conducted on the weeks 53 to 78 of the dataset (out-of-sample). The weeks 1 to 30 are used to initialise acc_i^{norm}.

Table 8: On-time service levels per ADI accuracy error and number of customer segments

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
$\overline{err}^{acc}$	**0.03**	**0.19**	**0.27**	**0.31**	**0.41**	**0.38**	**0.40**	**0.44**	**0.51**	**0.65**
|K| = 2	23%	31%	35%	22%	31%	25%	28%	29%	23%	28%
|K| = 4	24%	33%	35%	24%	31%	26%	32%	32%	25%	35%
|K| = 6	28%	34%	39%	25%	35%	29%	33%	34%	28%	**37%**
|K| = |I|	**35%**	**41%**	**43%**	**29%**	**42%**	**33%**	**37%**	**37%**	**29%**	31%

Table 7 presents the $\overline{err}^{acc}$ values calculated on the in-sample and the out-of-sample data. For confidentiality reasons, the real profitability of the customers is not provided in the used dataset. However, information on the relation of profitabilities of the customers within the dataset is available. For the numerical study, the per-piece profitabilities of the most and least profitable customers are set to 0.1€ and 0.067€, which represents a realistic relation between customer profitabilities in the semiconductor industry. Further, the level of supply shortage is set to 20%, i.e. the total demand of all customers exceeds the available ATP for every product by 20%. To generate the supply values, the average cumulated demand of always five periods is calculated and multiplied with 0.8. The result is taken as the ATP supply in the respective weeks.

The numerical study is implemented in Java. IBM ILOG CPLEX V12.6.0 is used to solve the customer segmentation, allocation planning, and order promising models. The study was performed on a personal computer with an Intel Xeon E7-4860 v2 processor with 2.6 GHz and 32GB RAM on a 64-bit Microsoft Windows 7 installation.

7.3.2 Framework parametrisation

Before the influence of the design of the customer contract portfolio on the demand fulfilment performance can be investigated, the demand fulfilment framework has to be parametrised. For this, a full factorial design of experiments is used on the first 52 weeks of the dataset. The customer scores are initialised on the first 30 weeks and the demand fulfilment framework is run on the weeks 31 to 52. For the number of customer segments, the values $\{2; 4; 6; |I|\}$ are used for $|K|$. The upward nesting level is varied between 0 and $|I|$ and the relative weights of α and ω, i.e. $\frac{\alpha}{(\alpha+\pi+\omega)}$ and $\frac{\omega}{(\alpha+\pi+\omega)}$, between 0 and 1 in ten equidistant steps.

Table 8 presents the OTSL resulting for the different values of $|K|$ at the service level-optimal relative weights of α and ω and an upward nesting level of 0. The table shows that, except for one portfolio, the service level-optimal number of customer segments equals the number of customers. Only for P10 showing exceptionally low ($\overline{err}^{acc} = 0.65$) predictive quality of demand data it is better to make use of aggregation. However, the size of these segments lies between two and three customers, which is much smaller than the segment size used in conventional demand fulfilment approaches (e.g. Meyr 2009).

If upward nesting is allowed (u>0), then the results are even more in favour of not combining customers into segments. This is because, upward nesting reduces the advantage of segmentation by adding flexibility in order promising also for allocation planning on customer-individual level.

The maximum OTSL in Table 8 for all contract portfolios is approximately 40%. These rather low values result from the experimental design. The distribution of ATP is not aligned with the

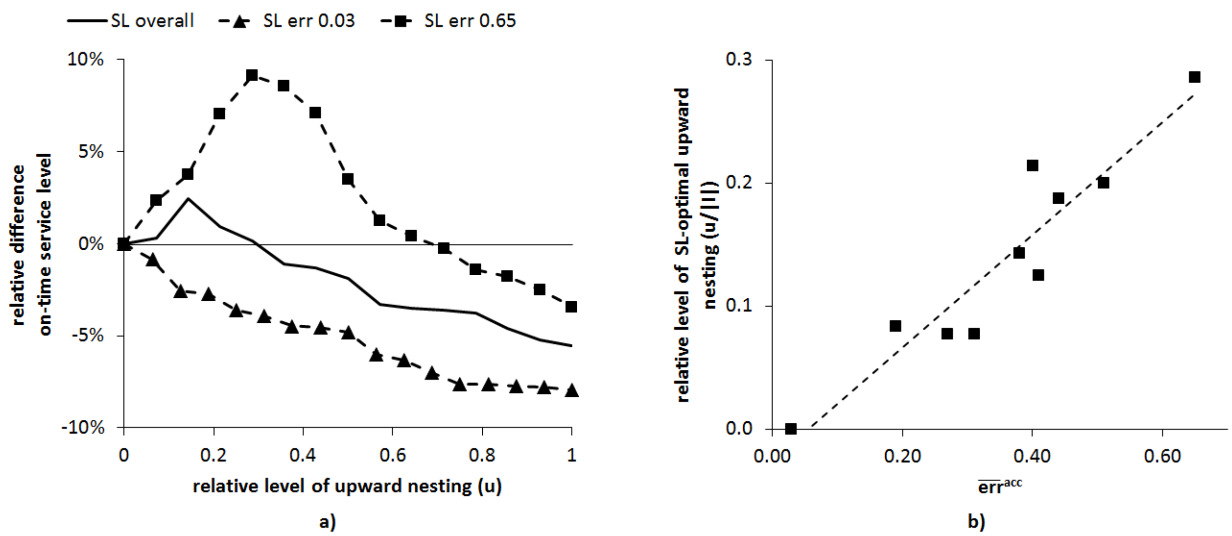

Figure 22: a) Service level by upward nesting level, b) service level-optimal upward nesting level by error of ADI accuracy.

distribution of the requested delivery dates of customer orders, the customer segments are nested, and the total customer demand exceeds the total available supply significantly by 20%. Consequently, early incoming orders consume large parts of the supply becoming available in later time periods. Hence, most of the later incoming orders cannot be promised on time. The theoretical maximum of 80% would be achieved, if all supply would be available already at the beginning of the experiment and the ADI and, hence, the AATP would be unbiased.

Figure 22 a) shows the average OTSL over all contract portfolios and the OTSLs of the contract portfolios with an $\overline{err}^{acc}$-value of 0.03 and 0.65 for different levels of $\frac{u}{|I|}$. All figures are shown relative to the OTSL at an upward nesting level of $u = 0$ and at the optimal levels of α and ω.

For all contract portfolios, the OTSL increases up to a certain level of u and decreases monotonically for higher values of u. For high u the OTSL falls below the OTSL without upward nesting. The reason for the increase of the service level for small u-levels is that errors in customer accuracy values can be compensated in the order promising process. However, with further increasing u, the order promising moves towards a FCFS approach since incoming orders can consume allocated supply quantities of any customer. It is shown in Chapter 6 that an FCFS order promising leads to lower service levels than allocation planning based demand fulfilment.

Figure 22 b) presents the service level-optimal upward nesting level $\frac{u}{|I|}$ for the $\overline{err}^{acc}$-values of the contract portfolios investigated. It shows that the choice of u depends on the predictive quality of data. For low values of $\overline{err}^{acc}$ the optimal u is smaller than for high values.

For the service-level-optimal values of u and $|K|$, Figure 23 shows the average relative difference of the OTSL relative to the case when α and ω are set to 0 for different levels of α and ω. The average is taken over all ten investigated portfolios. The figure illustrates that, for the in-sample data, the OTSL can be increased significantly (44% at the maximum) when customer order lead times and the accuracy of ADI of the customers are taken into account. In Section 7.3.3, it is investigated, if such benefits also exist for out-of-sample data.

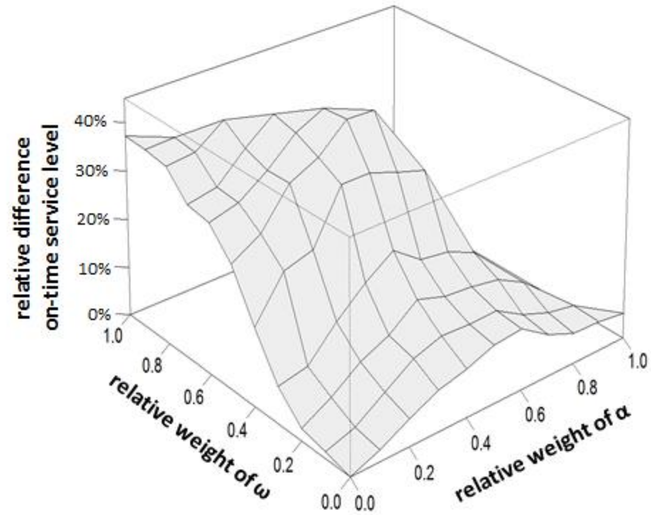

Figure 23: Average on-time service level performance of contract portfolios

Table 9 shows the service-level-optimal α, ω, and π levels for all ten contract portfolios separately. It shows that with decreasing predictive quality of ADI data customer order lead times get emphasized more in the customer scores. I.e. the service-level-optimal ω increases with increasing $\overline{err}^{acc}$. Therefore, it is important to know the predictive quality of data in order to be able to determine the correct values of α, ω, and π.

The impact of the parameter variations on profit is small. Compared to a profit-optimal parametrisation, the profit only declines by an average of 0.19%, when using the service-level-optimal parameters. Therefore, for the out-of-sample numerical experiments, the service-level-optimal parameters are used.

7.3.3 Numerical results

In the following, the demand fulfilment performance of the framework presented in Section 7.3.2 is investigated and insights on how to design contract portfolios for the set of customers are derived. For this, the parameters determined in Section 7.3.2 are used and the framework is run using the last 26 weeks of data in the datasets presented in Section 7.3.1.

7.3.3.1 Impact of parametrisation

In this section, the importance of exact parametrisation of the framework is determined and conclusions on its ease of implementation are drawn.

Table 10 shows the OTSL resulting from different levels of exact parametrisation. The second row of the table shows, which parameters are set to the exact values determined in Section 7.3.2. The parameters that are not shown in row two are set to the average of the exact values over all ten datasets.The table illustrates that exact parametrisation consistently leads to the highest OTSL. Comparing the values to the in-sample results in Table 8 shows that the usage of the framework on out-of-sample data leads to equally good results. I.e. the parametrisation of the approach robustly leads to good results also for out-of-sample data.

Table 10 further shows that using average values for the parameters generally results in good performance of the approach. This makes implementation easy. Exact parametrisation

Table 9: Service-level-optimal weight factors

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
$\overline{err}^{acc}$	**0.03**	**0.19**	**0.27**	**0.31**	**0.41**	**0.38**	**0.40**	**0.44**	**0.51**	**0.65**
α	0.6	0.5	0.4	0.5	0.4	0.4	0.3	0.2	0.1	0.0
ω	0.4	0.4	0.6	0.4	0.5	0.6	0.7	0.8	0.8	1.0
π	0.0	0.1	0.0	0.1	0.1	0.0	0.0	0.0	0.1	0.0

Table 10: Impact of exact parametrisation

	OTSL					
	-	$\boldsymbol{\alpha, \omega, \pi}$	$\boldsymbol{\alpha, \omega, \pi, \lvert K\rvert}$	$\boldsymbol{\lvert K\rvert, u}$	$\boldsymbol{\alpha, \omega, \pi, u}$	$\boldsymbol{\alpha, \omega, \pi, u, \lvert K\rvert}$
P1	32.2%	34.7%	34.7%	34.2%	35.9%	35.9%
P2	40.2%	41.0%	41.0%	40.6%	41.8%	41.8%
P3	43.2%	43.2%	43.2%	43.9%	43.9%	43.9%
P4	31.6%	33.1%	33.1%	31.6%	33.1%	33.1%
P5	42.9%	43.7%	43.7%	42.9%	43.7%	43.7%
P6	40.3%	40.3%	40.3%	40.3%	40.3%	40.3%
P7	36.4%	38.2%	38.2%	36.4%	38.2%	38.2%
P8	38.7%	40.7%	40.7%	40.9%	41.3%	41.3%
P9	34.0%	34.5%	34.5%	34.7%	35.4%	35.4%
P10	29.7%	32.3%	34.5%	35.4%	33.2%	36.2%
average	36.9%	38.2%	38.4%	38.1%	38.7%	39.0%

Table 11: The value of considering order lead time, ADI accuracy and profitability in demand fulfilment

	OTSL		
	$\alpha = 1$	$\omega = 1$	all
P1	28.7%	35.1%	35.9%
P2	29.9%	37.3%	41.8%
P3	31.2%	38.5%	43.9%
P4	25.9%	32.1%	33.1%
P5	30.5%	39.8%	43.7%
P6	28.2%	37.0%	40.3%
P7	28.5%	36.6%	38.2%
P8	27.1%	37.2%	41.3%
P9	26.4%	35.0%	35.4%
P10	26.1%	35.8%	36.2%

further improves the performance of the approach. The difference between the OTSL resulting from exact parametrisation and the OTSL resulting from parametrisation with averages can however be rather large (as, e.g., for P10). Depending on the portfolio, there is a significant contribution of exact parametrisation of all parameters.

7.3.3.2 Performance analysis of proposed demand fulfilment methodology

In this section, the performance of the framework is measured and the value of considering order lead times and ADI accuracy of customers is determined.

Table 11 shows the OTSL of the framework, when only ADI accuracy is considered ($\alpha = 1$), only order lead times are considered ($\omega = 1$), and when all parameters are taken into account (all). For all scenarios, the values of the parameters u and $|K|$ are set to the service-level-optimal values determined in Section 7.3.2.

Table 11 shows that considering all parameters of the framework consistently leads to the highest OTSL values. Only considering ADI accuracy leads to the lowest OTSL for all ten portfolios. Taking only order lead times into account leads to a significant increase of OTSL when compared to only considering ADI accuracy. For the portfolios showing low predictive quality of data, only considering order lead times results in good OTSL values also compared to the case making use of all parameters.

The analysis shows that taking order lead times into account has a high positive influence on the performance of the demand fulfilment process. This is because it allows prioritising customers with long order lead times in the allocation planning step. When their orders realise, excess allocations resulting from biased ADI can be redistributed without loss of OTSL because other customers place their orders later. The additional consideration of ADI accuracy reduces the risk of excess allocations and leads to even higher performance, if the predictive quality of historical ADI accuracy values is high enough.

7.3.3.3 Dependence of customer service levels on the design of contract portfolios

This section draws conclusions on how contract portfolios should be designed such that the resulting interdependencies of the length of customer order lead times and the accuracy of ADI maximise the overall OTSL of the supplier. For the analysis, the demand fulfilment framework is run on the last 26 weeks of the dataset, using the parametrisation determined in Section 7.3.2.

Figure 24 a) shows the OTSL of the ten portfolios in dependence of the average customer order lead time of the portfolios. The linear regression function evaluating the strength of the correlation of the length of the average customer order lead times and the resulting OTSL is shown as a dotted line. Its R^2 value of 0.00 shows that the two measures do not correlate. This is especially interesting because practitioners often solely focus on the negotiation of long order lead times with their customers. Figure 24 a), however, demonstrates that the length of the order lead time alone does not determine the OTSL.

Figure 24 b) shows the OTSL of the ten portfolios in dependence of the correlation $r(olt_i, acc_i)$ between order lead times and ADI accuracy. If $r(olt_i, acc_i)$ is positive, customers with long order lead times provide more accurate ADI than customers with shorter order lead times and vice versa.

Figure 24 b) shows better OTSL when order lead times and ADI accuracy are negatively correlated. Further, the OTSL grows with the strength of the correlation. The linear regression

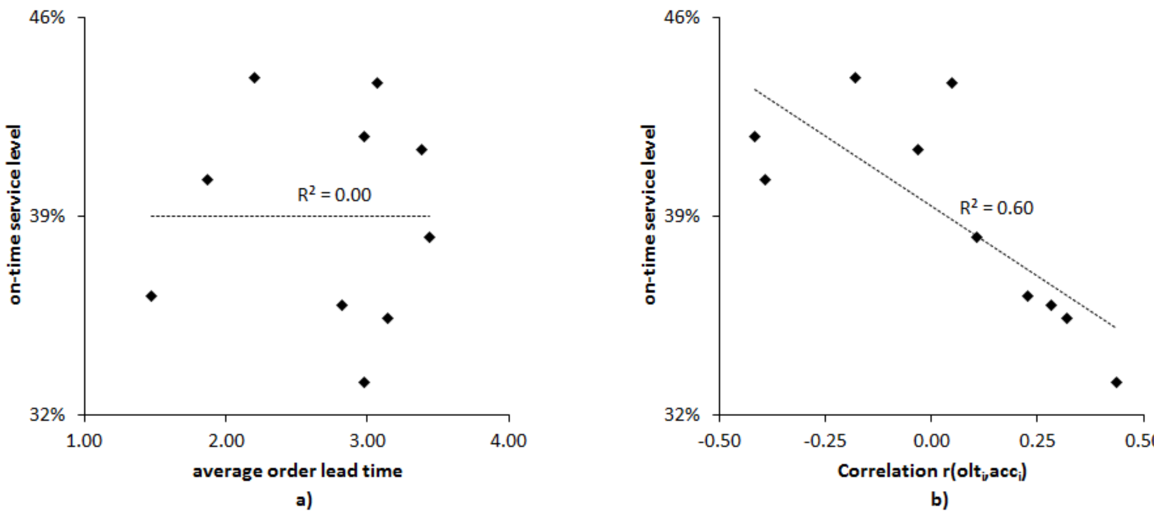

Figure 24: on-time service level in dependence of a) the average customer order lead time; b) the correlation of order lead time and advance demand information bias

function (dotted line) illustrates this negative correlation of $r(olt_i, acc_i)$ and OTSL. Its R^2 value of 0.60 indicates that it describes the relation well.

A negative $r(olt_i, acc_i)$ leads to high OTSL values because ADI-bias-caused excess allocations for customers with long order lead times can be reallocated to customers with shorter order lead times but more accurate ADI. The redistributed supply allocations are less likely to be excessive because the ADI of the customers to which the supply is allocated is more accurate. Hence, the utilization of the available supply as well as the overall OTSL are improved.

Figure 25 a) shows the portfolios' OTSL in dependence of the total demand share of customers with a high absolute ADI accuracy (acc_i) and a short relative order lead time (olt_i^{norm}) in the portfolio. The customers taken into consideration show an accuracy above 90% and a and olt_i^{norm} value of below 0.25. The demand share is calculated using realised demands. The R^2 value representing the strength of the correlation of the demand share and the OTSL shows that the volume of the demands being forecasted with a high absolute accuracy is not determining the OTSL performance of the portfolio. Figure 25 b) shows that it is more important for the supplier to create contract portfolios that lead to a significant demand share of customers showing high relative ADI accuracy (acc_i^{norm}) compared to the other customers in the portfolio. However, the positive correlation of the demand volume of customers with high relative accuracy and short relative order lead times and the OTSL performance of the portfolios, which is expressed in the R^2 value of 0.30, is not strong either.

Figure 26 a) and Figure 26 b) depict the OTSL in dependence of the share of total demands of customers with long relative order lead times (olt_i^{norm}) and low absolute (acc_i) and relative (acc_i^{norm}) ADI accuracy. In analogy to Figure 25 it is illustrated that there is no correlation between the OTSL resulting from the setup of a customer contract portfolio and the demand volume of customers showing low absolute ADI accuracy and relatively long order lead times, while there is a weak positive correlation of the OTSL performance of a portfolio and the demand share of customers with low relative ADI accuracy and long order lead times.

Summarizing the findings of the Figures 24, 25 and 26, it is more important to design contract portfolios such that order lead times and ADI accuracy are negatively correlated than

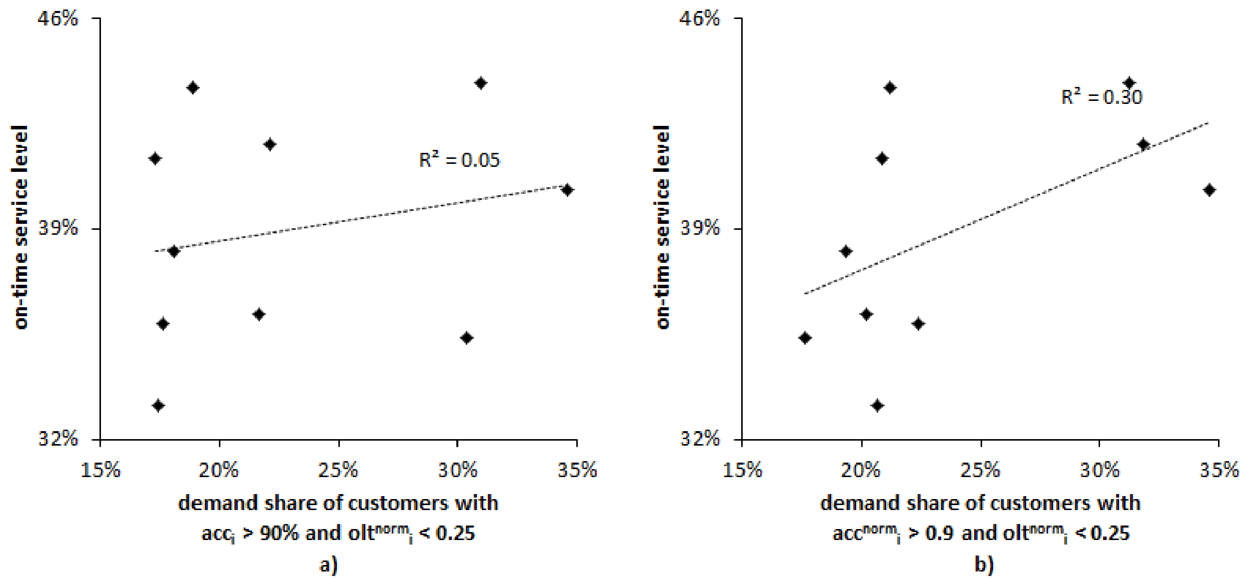

Figure 25: on-time service level in dependence of the demand share of customers with low relative order lead times and a) high absolute ADI accuracy and b) high relative ADI accuracy

to maximise the average order lead times of customers or to contract all customers for low (absolute) ADI accuracy. Customer contract portfolios have to be built such that there is a gradient of order lead times and ADI accuracies in the portfolio (see Figures 25 b) and 26 b)) and the correlation of the two measures is negative. Then, it is best possible to redistribute excess supply allocations to customers with long order lead times but low ADI accuracy to customers with short order lead times and high ADI accuracy.

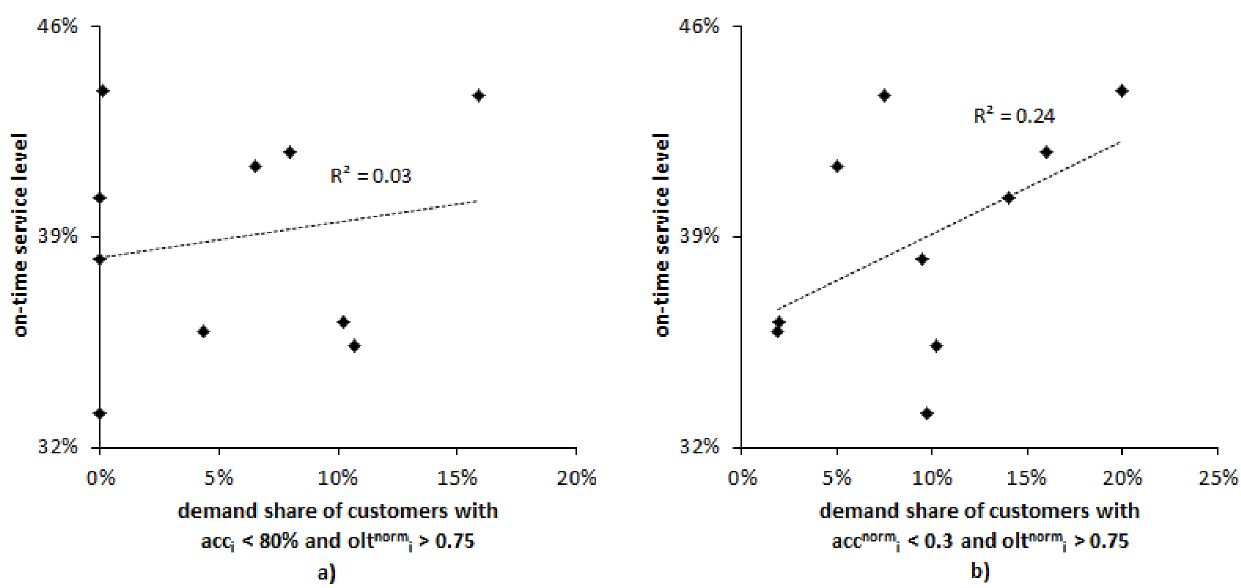

Figure 26: on-time service level in dependence of the demand share of customers with long relative order lead times and a) low absolute ADI accuracy and b) low relative ADI accuracy

7.4 Managerial implications and conclusion

When supply is short, supply distributed to customers by means of customer segmentation, supply allocation planning, and order promising. In this research, industrial contract portfolios with customer-specific terms for order lead times and ADI are analysed in order to derive insights for contract portfolio management. For this purpose, different portfolio designs are investigated in the dynamic context of industrial planning processes, for which a framework is developed that captures the interrelationship between order lead time and ADI. The approach extends the work presented in Chapter 6 by not only taking profitability and accuracy of ADI of individual customer into account, but also customer order lead times. In order to reflect changing customer ordering behaviour, a dynamic process is used to continuously update customer data. If the ordering behaviour of individual customers is hard to measure accurately, either upward nesting is allowed in order promising to increase the flexibility of the approach or customers are grouped into segments.

In a numerical study, it is first demonstrated how to set up the framework for different contract portfolios from the semiconductor industry. It is shown that demand fulfilment should take all contract terms, including order lead times and historical forecasting and ordering behaviour of all customers into account, that it should be performed on the individual customer level, and that it is of special importance to determine the predictive quality of ADI in order to be able to parametrise the framework right. Furthermore, the analysis illustrates that the framework leads to significant improvements in service levels and robustness in performance.

Second, insights aiding suppliers in their contract portfolio management are derived. The analysis shows that demand fulfillment performance is not primarily determined by the absolute length of the order lead times or the absolute level of ADI accuracy of the customers in the portfolio. Instead, the presence of a negative correlation with the accuracy of advance demand information in the entire contract portfolio is important. Then, excess allocations can be redistributed to other customers without loss of service levels. Consequently, suppliers must consider the portfolio of all customers and negotiate relatively long order lead times for customers showing relatively low accuracy of advance demand information. On the other hand, contracts with customers showing relatively high forecast accuracy should allow the customers to place their orders with short lead times.

For future research, considering additional data in demand fulfilment activities is interesting. For example, substitution of products taking into account the individual willingness of customers to substitute has not been dealt with so far. Also, including uncertainty of supply and volatility of demand after order realization is a further extension possibility. Moreover, further investigating the interactions between supply network planning, demand fulfilment, and customer contracting is an interesting direction of further research. In many industries, production quotas for supply planning are negotiated between different business divisions of a company. For example, efforts could be spent on integrating this type of allocation with supply allocation for customers taking contracts into account and streamlining all these activities in a holistic approach aiming at the maximisation of customer service levels and profits.

8 The value of data for demand fulfilment

This chapter summarizes and discusses the results of the research presented in this thesis (Section 8.1). The main scientific contributions are highlighted in Section 8.2. Managerial insights are derived in Section 8.3. Finally, the limitations of the presented research are detailed (Section 8.4). From these, directions for future research are derived in Section 8.5.

8.1 Summary and discussion of results

Taking the perspective of an industrial supplier, this thesis deals with the exploitation of big data in demand fulfilment related supply chain planning processes to increase the robustness and accuracy of demand fulfilment and raise customer service levels. The studied system consists of the processes demand planning and supply network planning, the demand fulfilment processes customer segmentation, allocation planning, and order promising, the interface of these processes to the set of customers and the contract portfolio a supplier has with their customers (see Section 1.2). The presented research focusses on the interdependencies between the processes and the data exchanged in the system.

For this purpose, Chapter 2 discusses the role of demand fulfilment in supply chain planning. Then, relevant literature from related disciplines is reviewed in Chapter 3. In Chapter 4, a data driven supply chain planning framework for robust and accurate demand fulfilment is presented. Chapters 5 to 7 detail parts of this framework and develop methods that exploit big data in order to increase demand fulfilment performance. In particular, a method that represents supply chain flexibilities in supply information used for order promising is developed in Chapter 5. In Chapter 6, an allocation planning approach considering the historical bias of ADI, i.e. demand forecasts, provided by the customers is proposed. Chapter 7 analyses industrial contract portfolios with customer-individual terms for order lead times and ADI in order to derive insights for portfolio management. For this purpose, the demand fulfilment methodology developed in Chapter 6 is extended to also consider data on the OLT of customers. Insights on the management of the portfolio of customer contracts are derived considering the dynamics of real-world demand fulfilment processes.

Five research questions are addressed, which are answered in the following considering the results of the research presented in this thesis.

RQ1. How should the supply chain planning processes of the studied system (Figure 1) be integrated in order to increase customer service levels and the robustness and accuracy of demand fulfilment?

The data driven supply chain planning framework developed in Chapter 4 shows that, in order to increase customer service levels and improve the robustness and accuracy of demand fulfilment, all planning decisions of the system need to be streamlined and consider demand fulfilment relevant supply chain and customer data. Enabler processes need to be integrated into the supply chain planning landscape, which provide the planning processes with data on the capabilities of the supply chain in terms of flexibility, the customer forecasting and ordering behaviour in terms of the accuracy of ADI and the length of OLTs, and the contractual, strategic, and operational obligations of the supplier towards its customers.

The concepts of robust and flexible planning have to be employed on different levels of the planning hierarchy in order to be able to react on short-term demand variations and utilise supply chain resources like capacities, materials, and supply efficiently. Flexibility is attained by steering incoming demand to the most profitable point in time using price-based revenue management methodologies in the order promising process. Robustness is achieved by improving the long-term, mid-term, and short-term demand planning processes, the supply network planning process, the customer segmentation process, and the allocation planning process by integrating the additional data mentioned above. Because several demand fulfilment related supply chain planning processes consider customer demand data directly in their decisions, customers must be incentivised to provide truthful ADI and long OLTs in order to increase the accuracy of planning. How this can be done is described in the answer to Research Question RQ4.

The research results presented in this thesis show that the most important factor for increasing service levels and the robustness and accuracy of demand fulfilment is the availability of highly granular data provided by the mentioned enabler processes to all planning processes in the system. Moreover, the allocation planning and order promising processes have to be provided with supply information reflecting the production capabilities of the supply chain, which go beyond planned supply for forecasted demands.

RQ2. How should available data on supply chain capabilities be considered in demand fulfilment in order to increase the accuracy and robustness of order promises in industrial settings with heterogeneous customer OLTs and uncertainty regarding the realisation of aggregated demand forecasts on finished product level?

Chapter 5 shows that representing supply chain flexibilities in the availability information used for order promising increases both the robustness and the accuracy of delivery date confirmations. More precisely, these flexibilities must be represented in the ATP information, which is calculated on basis of the master production schedule and used in the order promising process. It is shown that CTP methods, which use the supply network planning process to promise orders, have an even better capability of considering flexibilities in the supply chain and therefore increase the robustness and accuracy of order promises more than ATP based order promising approaches. However, CTP is usually not suitable for real-time order promising in industrial settings, because the necessary data and models are too complex and, hence, the computation times are prohibitively long.

The supply chain capabilities, which have to be represented in the ATP information, are product and process flexibilities. Product flexibility is the possibility to produce different products from one type of material. Process flexibility is the possibility to use one production process for the production of different products. These flexibilities can be presented in ATP information using the methodologies introduced in Chapter 5, which exploit data on supply chain flexibilities such as production cycle times, resource consumption factors, and BOM information and cumulate ATP quantities accordingly. The data has to be provided on individual product and process level. Then, changes in the master production schedule due to newly arrived orders can be anticipated and the accuracy and robustness of order promises can be increased.

The numerical analysis in Chapter 5 shows that the developed methodologies reduce the amount of orders receiving an initial order promise later than the earliest feasible delivery date. Thus, representing supply chain flexibilities in supply information increases the order winning probability of a company, because customers more often receive the earliest possible delivery date for their order. Furthermore, it is shown that the benefits of such approaches grow with increasing customer OLTs and degrees of flexibility in the supply chain.

RQ3. How should available data on the forecasting and ordering behaviour of customers be considered in demand fulfilment in order to increase service levels in supply shortage situations?

In supply shortage situations, the relevant interface of the supply chain planning processes with the ADI provided by the customers moves from supply network planning to allocation planning (see Chapter 7). Allocation planning uses ADI to allocate supply to customer segments. Since ADI, by its nature, is uncertain and possibly biased, the supply allocations as well as their consumption policy in order promising influence the service levels with which customers are served. Consequently, data on the forecasting and ordering behaviour of customers has to be considered in the demand fulfilment processes customer segmentation, allocation planning, and order promising.

Chapters 6 and 7 show that this data should be considered in the customer scores, which indicate the priority of the customers for demand fulfilment. Furthermore, the data should be considered on the individual customer and product level, if the predictive quality of historical ADI accuracy is not very low. If this is done, it is most beneficial to also perform allocation planning and order promising on the individual customer and product level. Furthermore, service levels can be increased by applying flexible concepts in demand fulfilments that compensate for low predictive quality of the used data. For example, upward nesting can be allowed in order promising or customers can be grouped into segments dynamically in order to pool uncertainties.

The studies show that, since the accuracy of ADI and the length of OLT interact in the demand fulfilment processes, both need to be taken into account. Then, the service levels of a supplier can be increased significantly, because supply can be allocated and consumed effectively and efficiently, leading to reduced stock levels that result from excess allocations to certain customer segments. Additionally, considering ADI and OLT data in demand fulfilment helps industrial suppliers to further improve the accuracy and robustness of their order promises, since re-allocations of supply become necessary less often.

The analyses further show that the benefits of considering customer forecasting and ordering behaviour data in demand fulfilment become more distinct with growing scarcity of supply and increased predictive quality of the available data. It is also demonstrated that the consideration of the accuracy of ADI improves service levels by raising the efficiency of allocation planning while the consideration of the length of OLT enables the supplier to raise service levels by effectively reallocating excess allocations.

RQ4. How does integrating data on customer forecasting and ordering behaviour into demand fulfilment processes increase planning security?

Customers systematically inflate their demand forecasts in supply shortage situations to make suppliers increase the supply quantities reserved for them and finally satisfy the total of their actual demand. This behaviour, called rationing gaming, reduces planning security since it is significantly impairing the ability of the supplier to efficiently allocate current and future supply to customers (see Chapter 6). Therefore, suppliers want their customers to provide orders with long OLTs. However, customers are usually not willing to place orders long in advance of their due date (see Chapter 7).

The study conducted in Chapter 6 demonstrates that customers can be incentivised to provide truthful ADI, when data on the historical bias of ADI is taken into account in the demand fulfilment processes of a supplier. To realise these incentives, customers providing ADI with low biases must be prioritised in demand fulfilment and the mechanisms of such demand fulfilment approaches must be communicated to customers. Because of the prioritisation, the service levels for customers with high ADI accuracy increase over-proportionally compared to the service levels of other customers. Furthermore, Chapters 6 shows that it is not a dominant strategy for the most profitable customers to still inflate their ADI systematically. Hence, the incentive to forecast their demands truthfully in ADI exists for all customers.

This work does not study the effect of the consideration of OLTs in demand fulfilment on the incentive for customers to provide their orders with longer OLTs. It can however be assumed that taking OLTs into account leads to a similar effect as the one demonstrated for ADI accuracy in Chapter 6, because considering OLTs enable the supplier to prioritise customers with long order lead times in demand fulfilment.

Consequently, the approaches developed in Chapters 6 and 7 lead to higher planning security and potentially enable the supplier to move the order decoupling point upstream, further revealing cost saving potential and flexibility that result from holding stocks of materials rather than final products.

RQ5. How should the portfolio of contractual agreements with the entire set of customers be managed in order to maximise the demand fulfilment performance of the supplier?

Chapter 7 shows that, since the accuracy of ADI and the length of customer OLTs interact in the demand fulfilment processes, the conditions for both parameters in form of maximum and minimum bounds for the deviation of ADI from the final order and minimum OLTs influence the overall demand fulfilment performance of the supplier. These conditions are determined in dyadic contracts negotiated between the supplier and its customers individually. Chapter 7 illustrates that, when negotiating new contracts, the entire portfolio of contracts the supplier holds with its customers must be considered in order to improve the supplier's demand fulfilment performance.

The analysis in Chapter 7 shows that only focussing on negotiating long OLTs with customers does not increase the performance of demand fulfilment. Instead, the contractual terms need to be designed in a way such that the length of OLTs and the accuracy of ADI are negatively correlated; i.e. customers providing their ADI with low accuracy have to be contracted for long OLTs and vice versa. Then, excess allocations for customers with low ADI accuracy can be redistributed without loss of service level.

8.2 Contributions

The research presented in this thesis contributes in several ways to the current state of the art. Chapter 4 presents a data driven framework for robust and accurate demand fulfilment in industrial environments. It consists of robust and flexible approaches for demand steering and dynamic pricing, extending current industry practice in several aspects. The concept of availabilities and capabilities (A&C), as well as various planning processes and process enablers are introduced. Further, the framework is the first to transfer revenue management ideas to industrial supply chain planning. It streamlines the decisions of demand planning, supply network planning, allocation planning, and order promising towards robustness and accuracy of order confirmations and integrates data on supply chain capabilities, customer ordering behaviour, and the contractual, strategic, and operational obligations of the supplier towards its customers.

Chapter 5 develops an order promising methodology able to deal with demand uncertainty and heterogeneous customer OLTs in industrial environments by considering supply chain flexibilities in ATP processes commonly used in practice. The approach exploits product and process flexibilities, which are formalized for the first time. An increased accuracy and robustness of order promises is demonstrated in a numerical study of real-time order promising based on an industry case.

Chapter 6 introduces a new allocation planning approach that exploits increasingly available data on individual customers and products by allocating supply on a highly granular level at high planning frequencies. More specifically, the methodology considers the demand forecast bias of customers and thereby supports an efficient supply allocation. The approach also incentivizes the customers to communicate truthful ADI. Using the method in a numerical study based on an industrial case it is demonstrated that the proposed methodology leads to lower average stock levels and an increased overall service level, especially for customers with a low ADI bias. The analysis further shows that supply allocation on individual customer level is only valuable, if additional data such as ADI bias and OLT is also exploited on this granularity level.

Chapter 7 derives insights for contract portfolio management by analysing industrial contract portfolios with customer-individual terms for order lead times and ADI. A framework is developed that extends the demand fulfilment methodology presented in Chapter 6 to investigate contract portfolios in the dynamic context of industrial planning processes. It is demonstrated how to apply and parametrise the framework to contract portfolios from the semiconductor sector. In a numerical study, it is shown that customers should not be grouped into segments but receive allocations individually. The analysis illustrates that the consideration of order lead times results in significant improvements in service levels and performance robustness. The results further show that those portfolios, in which order lead times and the accuracy of ADI are negatively correlated, perform substantially better. Such portfolios are even superior to portfolios in which all customers have long order lead times. They allow a more efficient reallocation of excess allocations.

8.3 Managerial insights

The results of the research presented in this thesis have several implications for practitioners. First, they show that planners have to consider the entire system of planning processes and their decisions in order to achieve planning stability, i.e. robustness and accuracy of order promises, and increase the service levels for their customers. The decision of single planning steps must be seen in the context of the ultimate goal of supply chain planning, i.e. achieving a competitive advantage through superior customer service. This means that local optima for single planning steps are to be avoided in case they counteract global optima of demand fulfilment performance.

For example, from a supply network planning point of view it might seem optimal to only plan production for demand forecasts and existing orders in order to reduce production and stock holding cost to a necessary level. However, from a demand fulfilment point of view it is necessary to also plan production to a certain target utilisation level of capacities so that supply allocation and order promising processes are able to promise robust and accurate delivery dates for unforeseen incoming orders (see Chapter 4). It might also seem logical for customer relationship managers to negotiate OLTs and ADI flexibility boundaries that are as long and narrow as possible, respectively. However, as shown in Chapter 7, the whole portfolio of customers, their contracts and their forecasting and ordering behaviour need to be considered in individual contract negotiations in order to maximise overall service levels. This means that it is necessary to negotiate narrow ADI flexibility boundaries and short OLTs with some customers, while wide ADI flexibility boundaries and long OLTs need to be agreed upon with others.

In consequence, the requirements towards supply chain planners in terms of education and knowledge increase. Planners need to be able to understand the planning methods and decisions in all relevant planning steps, which are part of the demand fulfilment context of the planning process they are responsible for. Furthermore, the planning methods and decisions also need to be transparent for all planners. In industrial practice, this is often not the case, because responsibilities are distributed and often customized planning tools are purchased from internal or external service providers, which do not provide sufficient documentation of the developed solutions.

Moreover, in order to achieve robustness in demand fulfilment, the planning processes must consider the interrelations of the planning system and reflect the characteristics and current status of the supply chain appropriately. Feedback loops between the planning processes of demand fulfilment and the physical supply chain have to be installed in order to be able to adapt planning. For example, the information about early production starts due to pre-production must be forwarded to the demand fulfilment processes when ATP is cumulated in order to represent supply chain flexibilities based on production cycle time information (see Chapter 5).

The results of this thesis also show that enriching planning processes with available data improves the quality of the decision and the performance of the planning system, if the type of the new data is appropriate regarding the purpose and objective of the planning step, in which it is integrated. Given sufficient predictive quality of the data, the granularity level of the decision must be adapted to the granularity of the used data in order to achieve best results. For example, the demand fulfilment methods developed in Chapter 6 and 7 perform best, if allocation planning and order promising decisions are taken on the individual customer level, on which also the used data on customer forecasting and ordering behaviour is available.

Consequently, practitioners should invest in the necessary infrastructures to collect, store, process, and analyse demand fulfilment relevant data. With the newly available big data processes and tools, this is made possible. However, many companies in different industries still lack this infrastructure.

Finally, the results presented here demonstrate that it is possible for suppliers to incentivise customers to change their forecasting and ordering behaviour only by using certain demand fulfilment policies. Importantly, this can be done without forcing certain behaviour through contracts. In order to do so, not only the demand fulfilment methodologies used by the supplier need to be adapted. The used allocation planning and order promising policies, especially regarding the dependency of the priorities of customers on their forecasting and ordering behaviour, must be communicated to the customers in order to become effectual. This is especially beneficial in environments, where customers, due to their market power, cannot be forced by contracts to reduce gaming behaviour or provide orders with long lead times. Using the methods developed in Chapter 6 and 7 and communicating their demand fulfilment policies to customers incentivises the customers to change their behaviour on their own in order to be served with high service levels.

8.4 Limitations

Due to its scope, the research presented in this thesis has several limitations. First, since all numerical studies use data from the semiconductor industry and the framework presented in Chapter 4 has been developed for a semiconductor manufacturer, there is a clear focus on this industry. However, considering the special characteristics of semiconductor manufacturing, i.e. high volatility of demand, long production lead time, capital intensive manufacturing, and particularly short product life cycles, the developed approaches are likely to also be beneficial in other industries, which show similar characteristics and are typically less dynamic. Nonetheless, the work should be complemented by testing the applicability of the approaches in other industrial environments.

The numerical study in Chapter 6 shows several case specific trade-offs between profitability and service levels depending on the weight of ADI accuracy in the prioritisation of customers for demand fulfilment. These mainly exist because the most profitable customers in the investigated dataset, due to their strong position towards the supplier, do not provide their orders with long OLTs and their ADI with high accuracy. On the other hand, the numerical study shows that the developed data driven demand fulfilment methodology incentivises customers to change their forecasting and ordering behaviour to provide truthful ADI. Hence, if customers consequently change their ordering behaviour, the basis for the analysis changes as well.

The thesis does not investigate how the resulting change in customer ordering behaviour affects the performance of the developed demand fulfilment approaches. However, the effects can be foreseen without an extensive study. First, because the most profitable customers are incentivized to provide truthful ADI, the changing customer behaviour leads to increased profits, because these customers adjust their ordering behaviour to increase their priority in demand fulfilment. Second, the changed customer behaviour leads to higher planning security for the supplier, resulting in even higher efficiency of supply allocations, lower average stock levels, and higher service levels. Therefore, the benefits of employing the developed approaches further increase when customers change their behaviour.

The research presented in this thesis is based on assumptions that exclude certain characteristics of real world supply chains. For example, capacity and supply uncertainty is excluded in all performed case studies. Moreover, buffer stocks for materials and finished products and corresponding supply contingencies are not considered. These assumptions are made to isolate the effects of demand uncertainty and customer gaming behaviour on the performance of the developed demand fulfilment methodologies and prove their improvement potential compared to conventional approaches. In practice, demand uncertainties interact with supply uncertainties and buffer stocks. The effect of these interdependencies on the performance of the developed methodologies is not tested. However, it is intuitive to assume that the positive effects of the presented approaches on demand fulfilment performance remain under these conditions as well. Only their magnitude decreases.

Furthermore, in all studies, ADI and firm orders are clearly differentiated. In reality, logistic concepts exist, which do not allow such a clear distinction. For example, customers can change requested quantities and delivery dates within certain boundaries also for orders. Therefore, if the developed approaches are implemented in practice, they need to be modified to also consider uncertainties in orders. This can be done by including a respective term in the customer scoring approaches, allocate supply not only on basis of ADI but also on order and internal demand forecast information considering this additional uncertainty, and performing order re-promising frequently for all open orders in the system.

Chapter 7 lacks an analysis, if the presented demand fulfilment approach incentivises customers to provide their orders with long OLTs. It can however be assumed that taking OLTs into account leads to a similar effect as the one demonstrated for ADI accuracy in Chapter 6, because considering OLTs enables the supplier to prioritise customers with long order lead times in demand fulfilment. Nonetheless, investigating the effects of the demand fulfilment methodology presented in Chapter 7 on the incentive for customers to provide long OLTs would be a worthwhile extension of this work.

Finally, this thesis does not consider the roles and behaviour of human planners in supply chain planning processes. This is standard in the related literature. However, in practice, human planners might take sub-optimal or irrational decisions regarding the global demand fulfilment optimum because service levels for certain customers are maximised based on individual situations. Then, the performance of the developed approaches might differ from the analyses presented in this thesis.

8.5 Directions for future research

There are many possibilities to develop the research presented in this thesis further. Based on the framework presented in Chapter 4, first, new forecasting and optimisation techniques for demand planning can be developed, which include the mentioned internal and external sources of information. Additionally, the long- and mid-term demand planning and supply chain planning processes are currently predominantly done manually in the industry. Automated decision support systems for these processes would increase their efficiency and transparency substantially. Cloud and big data analysis tools and crowd opinion techniques are promising technologies to drive automation in these fields.

A rather managerial challenge is to establish machine-readable databases containing contract and customer forecasting and ordering behaviour data and integrate these into current ERP systems. The high practical efforts of establishing and maintaining such a database make organizations reluctant to implement such databases. Also, a method implementing the shown dynamic pricing and order confirmation process has to be developed. The challenge is to find industry-suitable models for customer price sensitivity as well as opportunity cost representations for supply consumption.

For the study conducted in Chapter 5, it is assumed that except demand there are no other sources of uncertainty. It would be interesting to study the dependency of the robustness of order promises on the accuracy of estimated cycle times when demand uncertainties interact with capacity and supply uncertainties and safety stocks. When utilization is fluctuating and cycle times are dynamic, the method needs to be extended to reflect additional information about the shop floor. Here, the consideration of order individual production cycle times as well as supply chain capabilities, i.e. unused capacities and swopping possibilities in the production schedule, are potential starting points.

To include more aspects of real world industrial supply chains, the research presented in Chapters 6 and 7 could be extended to consider substitute products in the allocation planning and order promising processes. Especially taking data on the individual willingness of customers to substitute products into account has so far not been addressed. The examination of the performance of the approaches under supply and demand uncertainty after order arrival is a further extension possibility. Here, order rescheduling and cancellation rules, which industrial suppliers and customers agree upon in supply contracts, have to be taken into consideration. Moreover, in many industries, production quotas for supply network planning are negotiated between different business divisions of a company. Efforts could be spent on integrating this type of allocation with supply allocation for customers taking contracts into account. Also, the benefits of the approach could be confirmed for changing forecasting and ordering behaviour in response to the set incentives for accurate forecasting and long term ordering.

As mentioned in Section 8.4, it is also interesting to test the developed approaches for industries other than the semiconductor sector in order to analyse their performance under different environmental characteristics and derive more improvement potential.

On a more general scale, further investigating the interactions between supply network planning, demand fulfilment, and customer contracting is an interesting direction of further research. In particular, new approaches need to be developed for environments, which do not show a clear distinction between firm orders and customer forecasts. Moreover, suitable performance indicators and tools need to be developed, which establish the necessary transparency for human planners in the supply chain to work towards global instead of local optima and fulfil customer requirements while increasing revenues at the same time. Simulation techniques combining discrete-event and agent-based modelling could help to achieve this goal since they are capable of modelling system as well as human behaviour. Such tools can also help to analyse the interactions of human planners, planning methods, and planning tools, like ERP systems, supply network planning, allocation planning, and order promising. With the results, misalignments of these planning elements, which cause instabilities in the system, can be identified.

References

Aberdeen Group, 2013. CSCO 2014: Top three supply chain execution priorities. Available at: http://v1.aberdeen.com/launch/report/perspective/8757-AI-supply-chain-priorities.asp. Accessed: 14.04.2016.

Adida, E., Perakis, G., 2010. Dynamic pricing and inventory control: robust vs. stochastic uncertainty models - a computational study. Annals of Operations Research, 181(1), 125-157.

Alarcón, F., Alemany, M. M. E., Ortiz, A., 2009. Conceptual framework for the characterization of the order promising process in a collaborative selling network context. International Journal of Production Economics, 120(1), 100-114.

Alemany, M., Grillo, H., Ortiz, A., Fuertes-Miquel, V., 2015. A fuzzy model for shortage planning under uncertainty due to lack of homogeneity in planned production lots. Applied Mathematical Modelling, 39(15), 1-35.

Albrecht, M., Rohde, J., Wagner, M., 2015. Master planning. In: Stadtler, H., Kilger, C., Meyr, H., (Eds.). Supply Chain Management and Advanced Planning. Berlin: Springer, 155-175.

Ali, M., Gaudreault, J., D'Amours, S., Carle, M.-A., 2014. A multi-level framework for demand fulfillment in a make-to-stock environment - a case study in Canadian softwood lumber industry. 10ème Conférence Francophone de Modélisation, Optimisation et Simulation, November 2014, Nancy, France.

Altendorfer, K., Minner, S., 2014. A comparison of make-to-stock and make-to-order in multi-product manufacturing systems with variable due dates. IIE Transactions, 46(3), 197-212.

Anderson, T. W., Darling, D. A., 1954. A test of goodness of fit. Journal of the American Statistical Association, 49(268), 765-769.

Armstrong, J., 1985. Long Range Forecasting: From Crystal Ball to Computer. New York: Wiley-Interscience.

Babarogić, S., Makajić-Nikolić, D., Lečić-Cvetković, D., Atanasov, N., 2012. Multi-period customer service level maximization under limited production capacity. International Journal of Computers Communications and Control, 7(5), 798-806.

Bakal, I., Erkip, N., Güllü, R., 2011. Value of supplier’s capacity information in a two-echelon supply chain. Annals of Operations Research, 191(1), 115-135.

Baker, K., Trietsch, D., 2015. Trading off due-date tightness and job tardiness in a basic scheduling model. Journal of Scheduling, 18(3), 305-309.

Balakrishnan, N., Sridharan, V., Patterson, J. W., 1996. Rationing capacity between two product classes. Decision Sciences, 27(2), 185-214.

Ball, M. O., Chen, C.-Y., Zhao, Z.-Y., 2003. Material compatibility constraints for make-to-order production planning. Operations Research Letters, 31(6), 420-428.

Ball, M. O., Chen, C.-Y., Zhao, Z.-Y., 2004. Available to promise. In: Simchi-Levi, D., Wu, D. D., Shen, Z.-J., (Eds.). Handbook of Quantitative Supply Chain Analysis. New York: Springer, 447-483.

Barnes-Schuster, D., Bassok, Y., Anupindi, R., 2006. Optimizing delivery lead time/inventory placement in a two-stage production/distribution system. European Journal of Operational Research, 174(3), 1664-1684.

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

Bassok, Y., Anupindi, R., Akella, R., 1999. Single-period multiproduct inventory models with substitution. Operations Research, 47(4), 632-642.

Bergeron, D., 2008. More than Moore. Proceedings of the 2008 IEEE Custom Integrated Circuits Conference (CICC), IEEE, San Jose, CA.

Brahimi, N., Aouam, T., Aghezzaf, E.-H, 2014. 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.

Buzacott, J.A., Shanthikumar, J. G., 1994. Safety stock versus safety time in MRP controlled production systems. Management Science, 40(12), 1678–1689.

Cachon, G. P., 2003. Supply chain coordination with contracts. In: de Kok, A.G., Graves, S.C., (Eds.). Handbooks in Operations Research and Management Science. Elsevier, 227-339.

Cachon, G. P., Lariviere, M. A., 1999a. Capacity allocation using past sales: When to turn-and-earn. Management Science, 45(5), 685-703.

Cachon, G. P., Lariviere, M. A., 1999b. Capacity choice and allocation: Strategic behavior and supply chain performance. Management Science, 45(8), 1091-1108.

Cederborg, O., Rudberg, M., 2009. Customer segmentation and capable-to-promise in a capacity constrained manufacturing environment. Proceedings of the 16th International Annual EurOMA Conference, Göteborg, Sweden.

Chatterjee, A., Cohen, M., Maxwell, W., Miller, L., 1984. Manufacturing flexibility: Models and measurements. Proceedings of the 1st ORSA/TIMS Special Interest Conference on FMS, Elsevier, Amsterdam, NL, 49-64.

Chen, M., 2006. Coordinating demand fulfillment with supply across a dynamic supply chain. Dissertation, University of Maryland.

Chen, J., Dong, M., 2014. Available-to-promise-based flexible order allocation in ATO supply chains. International Journal of Production Research, 52(22), 6717-6738.

Chen, A., Hsu, C.-H., Blue, J., 2007. Demand planning approaches to aggregating and forecasting interrelated demands for safety stock and backup capacity planning. International Journal of Production Research, 45(10), 2269-2294.

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

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

Chen, J., Chen, Y., Parlar, M., Xiao, Y., 2011. Optimal inventory and admission policies for drop-shipping retailers serving in-store and online customers. IIE Transactions, 43(5), 332-347.

Cheng, T. C. E., Gupta, M. C, 1989. Survey of scheduling research involving due date determination decisions. European Journal of Operational Research, 38(2), 156-166.

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

Chew, E., Lee, L., Liu, S., 2013. Dynamic rationing and ordering policies for multiple demand classes. OR spectrum, 35(1), 127-151.

Chiang, C., Hsu, H.-L., 2014. An order fulfillment model with periodic allocation review mechanism in semiconductor foundry plants. IEEE Transactions on Semiconductor Manufacturing, 27(4), 489-500.

Chiang, D. M.-H., Wu, A. W.-D., 2011. Discrete-order admission ATP model with joint effect of margin and order size in a MTO environment. International Journal of Production Economics, 133(2), 761-775.

Chien, C.-F., Wu, J.-Z., Wu, C.-C., 2013. A two-stage stochastic programming approach for new tape-out allocation decisions for demand fulfillment planning in semiconductor manufacturing. Flexible Services and Manufacturing Journal, 25(3), 286-309.

Christopher, M., 1998. Logistics and Supply Chain Management: Strategies for Reducing Cost and Improving Service. London: Prentice-Hall.

Claudio, D., Krishnamurthy, A., 2009. Kanban-based pull systems with advance demand information. International Journal of Production Research, 47(12), 3139-3160.

Costantino, F., Di Gravio, G., Shaban, A., Tronci, M., 2015. A real-time SPC inventory replenishment system to improve supply chain performances. Expert Systems with Applications, 42(3), 1665-1683.

Croxton, K., 2003. The order fulfillment process. The International Journal of Logistics Management, 14(1), 19-32.

Davizón, Y. A., de J. Lozoya, J., Soto, R., 2010. Demand management in semiconductor manufacturing: A dynamic pricing approach based on fast model predictive control. Proceedings of the 2010 IEEE Electronics, Robotics and Automotive Mechanics Conference (CERMA), IEEE, Cuernavaca, Mexico.

de Véricourt, F., Karaesmen, F., Dallery, Y., 2002. Optimal stock allocation for a capacitated supply system. Management Science, 48(11), 1486-1501.

Deshpande, V., Cohen, M., Donohue, K., 2003. A threshold inventory rationing policy for service-differentiated demand classes. Management Science, 49(6), 683-703.

Dickersbach, J. T., 2009. Supply Chain Management with APO: Structures, Modelling Approaches and Implementation of SAP SCM 2008. Berlin: Springer.

Dumitrescu, S., Steiner, G., Zhang, R., 2015. Optimal delivery time quotation in supply chains to minimize tardiness and delivery costs. Journal of Scheduling, 18(1), 3-13.

Ehm, H., Ponsignon, T., Kaufmann, T., 2011. The global supply chain is our new fab: Integration and automation challenges. Proceedings of the 22nd Advanced Semiconductor Manufacturing Conference (ASMC), IEEE, Saratoga Springs, NY, 1-6.

Enders, P., Adan, I., Scheller-Wolf, A., van Houtum, G.-J., 2014. Inventory rationing for a system with heterogeneous customer classes. Flexible Services and Manufacturing Journal, 26(3), 344-386.

Eppler, S., 2015. Allocation Planning for Demand Fulllment in Make-to-Stock Industries; A Stochastic Linear Programming Approach. Dissertation, Universität Darmstadt.

Ervolina, T. R., Ettl, M., Lee, Y. M., Peters, D. J., 2009. Managing product availability in an assemble-to-order supply chain with multiple customer segments. In: Günther, H.-O., Meyr, H., (Eds.). Supply Chain Planning. Berlin: Springer, 1-24.

Fildes, R., Kingsman, B., 2011. Incorporating demand uncertainty and forecast error in supply chain planning models. Journal of the Operational Research Society, 62(3), 483-500.

Fleischmann, B., Geier, S., 2012. Global available-to-promise. In: Stadtler, H., Fleischmann, B., Grunow, M., Meyr, H., Sürie, C., (Eds.). Advanced Planning in Supply Chains. Berlin: Springer, 195-215.

Fleischmann, B., Meyr, H., 2003. Planning hierarchy, modelling and advanced planning systems. In: de Kok, A. G., Graves, S. C., (Eds.). Handbooks in Operations Research and Management Science. Amsterdam: Elsevier, 457–523.

Fleischmann, B., Meyr, H., 2004. Customer orientation in advanced planning systems. In: Dyckhoff, H., Lackes, R., Reese, J., (Eds.). Supply Chain Management and Reverse Logistics. Berlin: Springer, 297-321.

Fleischmann, B., Meyr, H., Wagner, M., 2015. Advanced planning. In: Stadtler, H., Kilger, C., (Eds.). Supply Chain Management and Advanced Planning. Berlin: Springer, 71-95.

Framinan, J. M., Leisten, R., 2010. Available-to-promise (ATP) systems: A classification and framework for analysis. International Journal of Production Research, 48(11), 3079-3103.

Frank, K., Zhang, R., Duenyas, I., 2003. Optimal policies for inventory systems with priority demand classes. Operations Research, 51(6), 993-1002.

Gallego, G., Katircioglu, K., Ramachandran, B., 2006. Semiconductor inventory management with multiple grade parts and downgrading. Production Planning and Control, 17(7), 689-700.

Gallego, G., Stefanescu, C., 2009. Upgrades, upsells and pricing in revenue management. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1334341. Accessed: 21.03.2015.

Gallego, G., Talebian, M., 2012. Demand learning and dynamic pricing for multi-version products. Journal of Revenue and Pricing Management, 11(3), 303-318.

Gallego, G., Van Ryzin, G., 1994. Optimal dynamic pricing of inventories with stochastic demand over finite horizons. Management Science, 40(8), 999-1020.

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

Geier, S., 2014. Demand Fulfillment bei Assemble-to-Order-Fertigung. Produktion und Logistik. Wiesbaden: Springer.

Giard, V., Mendy, G., 2008. Scheduling coordination in a supply chain using advance demand information. Production Planning and Control 19(7), 655-667.

Gössinger, R., Kalkowski, S., 2015. Robust order promising with anticipated customer response. International Journal of Production Economics, 170, 529-542.

Gordon, V., Strusevich, V., Dolgui, A., 2012. Scheduling with due date assignment under special conditions on job processing. Journal of Scheduling, 15(4), 447-456.

Ha, A., 1997. Inventory rationing in a make-to-stock production system with several demand classes and lost sales. Management Science, 43(8), 1093-1103.

Han, G., Dong, M., Liu, S., 2014. Yield and allocation management in a continuous make-to-stock system with demand upgrade substitution. International Journal of Production Economics, 156(1), 124-131.

Hanke, J. E., Wichern, D. E., 2008. Business Forecasting. New Jersey: Pearson.

Hariharan, R., Zipkin, P., 1995. Customer-order information, leadtimes, and inventories. Management Science, 41(10), 1599-1607.

Harris, F., 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.

Huaili, C., Yanrong, N., 2010. Proceedings of the 2nd International Conference on Communication Systems, Networks and Applications (ICCSNA), IEEE, Hong Kong, China, 14-18.

Huang, K., 2005. Multi-stage stochastic programming models in production planning. Dissertation, Georgia Institute of Technology.

Huang, Y. S., Chen, J. M., Lin, Z. L., 2013. A study on coordination of capacity allocation for different types of contractual retailers. Decision Support Systems, 54(2), 919-928.

Huefner, R. J., Largay, J. A., 2013. Identifying revenue opportunities via capacity analysis. Journal of Revenue and Pricing Management, 12(4), 305-312.

Hung, Y.-F., Hsiao, J.-Y., 2013. Inventory rationing decision models during replenishment lead time. International Journal of Production Economics, 144(1), 290-300.

Hung, Y.-F., Lee, T.-Y., 2010. Capacity rationing decision procedures with order profit as a continuous random variable. International Journal of Production Economics, 125-136.

Hung, Y.-F., Tsai, P.-H., Wu, G.-H., 2014. Application extensions from the stochastic capacity rationing decision approach. International Journal of Production Research, 52(6), 1695-1710.

Ioannidis, S., 2011. An inventory and order admission control policy for production systems with two customer classes. International Journal of Production Economics, 131(2), 663-673.

Iyer, A., Bergen, M.E., 1997. Quick response in manufacturer-retailer channels. Management Science, 43(4), 559-570.

Janiak, A., Janiak, W. A., Krysiak, T., Kwiatkowski, T., 2015. A survey on scheduling problems with due windows. European Journal of Operational Research, 242(2), 347-357.

Jeong, B., Sim, S.-B., Jeong, H.-S., Kim, S.-W., 2002. An available-to-promise system for TFT LCD manufacturing in supply chain. Computers and Industrial Engineering, 43(1-2), 191-212.

Jin, X., Li, K., Sivakumar, A., 2013. Scheduling and optimal delivery time quotation for customers with time sensitive demand. International Journal of Production Economics, 145(1), 349-358.

Jung, H., 2010. An available-to-promise model considering customer priority and variance of penalty costs. The International Journal of Advanced Manufacturing Technology, 49(1-4), 369-377.

Jung, H., 2012. An available-to-promise process considering production and transportation uncertainties and multiple performance measures. International Journal of Production Research, 50(7), 1780-1798.

Kaminsky, P., Hochbaum, D., 2004. Due date quotation models and algorithms. In: Leung, J. Y.-T., (Eds.). Handbook of Scheduling: Algorithms, Models, and Performance Analysis, Boca Raton: CRC Press, 20-1-20-22.

Kaminsky, P., Kaya, O., 2006. MTO-MTS production systems in supply chains. University of California.

Kaminsky, P., Kaya, O., 2008. Scheduling and due-date quotation in a make-to-order supply chain. Naval Research Logistics, 55(5), 444-458.

Kaminsky, P., Kaya, O., 2009. Combined make-to-order/make-to-stock supply chains. IIE Transactions, 41(2), 103-119.

Kaminsky, P., Lee, Z., 2008. Effective on-line algorithms for reliable due date quotation and large-scale scheduling. Journal of Scheduling, 11(3), 187-204.

Karabuk, S., Wu, S. D., 2003. Coordinating strategic capacity planning in the semiconductor industry. Operations Research, 51(6), 839-849.

Karaesmen, F., Buzacott, J. A., Dallery, Y., 2002. Integrating advance order information in production control. IIE Transactions, 34(8), 649–662.

Karrer, C., Alicke, K., Günther, H.-O., 2012. A framework to engineer production control strategies and its application in electronics manufacturing. International Journal of Production Research, 50(22), 6595-6611.

Kilger, C., Meyr, H., 2015. Demand fulfillment and ATP. In: Stadtler, H., Kilger, C., Meyr, H., (Eds.). Supply Chain Management and Advanced Planning. Berlin: Springer, 177-194.

Kilger, C., Wagner, M., 2015. Demand planning. In: Stadtler, H., Kilger, C., (Eds.). Supply Chain Management and Advanced Planning, Berlin: Springer, 125-154.

Kim, W., 2011. Order quantity flexibility as a form of customer service in a supply chain contract model. Flexible Service and Manufacturing Journal, 23(1), 290–315.

Kim, J. K., Park, S. I., Shin, K. Y., 2014. A quantity flexibility contract model for a system with heterogeneous suppliers. Computers and Operations Research, 41(1), 98-108.

Klein, O., 2009. Fehlmengenverteilung im Demand Fulfillment. Göttingen: Cuvillier.

Knoblich, K., Heavey, C., Williams, P., 2015. Quantitative analysis of semiconductor supply chain contracts with order flexibility under demand uncertainty: A case study. Computers and Industrial Engineering, 87(1), 394-406.

Kremer, M., Van Wassenhove, L. N., 2014. Willingness to pay for shifting inventory risk: The role of contractual form. Production and Operations Management, 23(2), 239-252.

Leachman, R. C., 1993. Modeling techniques for automated production planning in the semiconductor industry. In: Ciriani, T. A., Leachman, R. C., (Eds.). Optimisation in Industry, Chichester: Wiley, 1-30.

Lebreton, B., 2015. Integrated campaign planning, scheduling and order confirmation in the specialty chemicals industry. In: Stadtler, H., Kilger, C., Meyr, H., (Eds.). Supply Chain Management and Advanced Planning. Berlin: Springer, 475-485.

Lečić-Cvetković, D., Atanasov, N., Babarogić, S., 2010. An algorithm for customer order fulfillment in a make-to-stock manufacturing system. International Journal of Computers, Communications and Control, 10(5), 783-791.

Lee, G., 2010. Real-time order promising methods considering scheduling of production lines under make-to-order environments. Proceeding of the 40th International Conference on Computers and Industrial Engineering (CIE), IEEE, Awaji, Japan, 1-5.

Lee, H. L., C. Billington, 1993. Material management in decentralized supply chains. Operations Research, 41(5), 835-847.

Lee, H. L., Padmanabhan, V., Whang, S., 2004. Information distortion in a supply chain: the bullwhip effect. Management Science, 50(12), 1875-1886.

Li, S., Ng, C., Yuan, J., 2011. Group scheduling and due date assignment on a single machine. International Journal of Production Economics, 130(2), 230-235.

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

Liu, S., Song, M., Tan, K., Zhang, C., 2015. Multi-class dynamic inventory rationing with stochastic demands and backordering. European Journal of Operational Research, 244(1), 153-163.

Lutze, H., Özer, Ö., 2008. Promised lead-time contracts under asymmetric information. Operations Research, 56(4), 898-915.

Makridakis, S. G., Wheelwright, S. C., Hyndman, R. J., 1998. Forecasting: Methods and Applications, New York: Wiley.

Massey, F. J., 1951. The Kolmogorov-Smirnov test for goodness of fit. Journal of the American Statistical Association, 46(253), 68-78.

Mentzer, J. T., Bienstock, C. C., 1998. Sales Forecasting Management, Thousand Oaks: SAGE Publications.

Meyr, H., 2008. Clustering methods for rationing limited resources. In: Mönch, L., Pankratz, G., (Eds.). Intelligente Systeme zur Entscheidungsunterstützung. Konferenzband zur Teilkonferenz der Multikonferenz Wirtschaftsinformatik München. San Diego: SCS Publishing House e.V., 19-31.

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

Meyr, H., 2012. Demand planning. In: Stadtler, H., Fleischmann, B., Grunow, M., Meyr, H., Sürie, C., (Eds.). Advanced Planning in Supply Chains. Berlin: Springer, 67-108.

Modarres, M., Zaefarian, T., Sharifyazdi, M., 2012. Stochastic capacity allocation, revenue management approach: the existence of modularity property. The International Journal of Advanced Manufacturing Technology, 60(5-8), 707-722.

Mönch, L., Fowler, J. W., Mason, S. J, 2013. Production Planning and Control for Wafer Fabrication Facilities: Modeling, Analysis, and Systems. New York: Springer

Moore, G. E., 1965. Cramming more components onto integrated circuits. Electronics, 38(8), 114-117.

Moses, S., Grant, H., Gruenwald, L., Pulat, S., 2004. Real-time due-date promising by build-to-order environments. International Journal of Production Research, 42(20), 4353-4375.

Nguyen, T.-H., Wright, M., 2014. Lead Time Quotation under Time-Varying Demand and Capacity. Bangkok: Atlantis Press.

Oracle, Capgemini, 2013. From customer orders through fulfillment: challenges and opportunities in manufacturing, high-tech and retail. Available at: https://www.capgemini.com /resource-file-access/resource/pdf/from_customer_orders_through_fulfillment_ study_english_final_0.pdf. Accessed: 30.10.2015.

Orlicky, J. A., 1975. Material Requirements Planning: The New Way of Life in Production and Inventory Management. New York: McGraw-Hill.

Ott, H. C., Heilmayer, S., Sng, C. S. Y., 2013. Granularity dependency of forecast accuracy in semiconductor industry. Research in Logistics and Production, 3(1), 49-58.

Ouyang, Y., 2014. Experimental study on using advance demand information to mitigate the bullwhip effect via decentralised negotiations. Transportmetrica B: Transport Dynamics, 2(3), 169-187.

Ouyang, Y., Daganzo, C. F., 2006. Counteracting the bullwhip effect with decentralized negotiations and advance demand information. Physica A, 363(1), 14-23.

Pang, Z., Shen, H., Cheng, T., 2014. Inventory rationing in a make-to-stock system with batch production and lost sales. Production and Operations Management, 23(7), 1243-1257.

Papier, F., 2016. Supply Allocation Under Sequential Advance Demand Information. Operations Research 64(2), 341-361.

Pibernik, R., 2005. Advanced available-to-promise: Classification, selected methods and requirements for operations and inventory management. International Journal of Production Economics, 93(1), 239-252.

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

Pibernik, R., Yadav, P., 2008. Dynamic capacity reservation and due date quoting in a make-to-order system. Naval Research Logistics (NRL), 55(7), 593-611.

Pibernik, R., Yadav, P., 2009. Inventory reservation and real-time order promising in a make-to-stock system. OR Spectrum, 31(1), 281-307.

Pinto, R., 2012. Stock rationing under service level constraints in a vertically integrated distribution system. International Journal of Production Economics, 136(1), 231-240.

Plambeck, E. L., Taylor, T. A., 2007. Implications of breach remedy and renegotiation design for innovation and capacity. Management Science, 53(12), 1859-1871.

Quante, R., Fleischmann, M., Meyr, H., 2009a. A Stochastic Dynamic Programming Approach to Revenue Management in a Make-to-Stock Production System, Rotterdam: ERIM.

Quante, R., Meyr, H., Fleischmann, M., 2009b. Revenue management and demand fulfillment: Matching applications, models, and software. OR Spectrum, 31(1), 31-62.

Rabbani, M., Monshi, M., Rafiei, H., 2014. A new AATP model with considering supply chain lead-times and resources and scheduling of the orders in flowshop production systems: A graph-theoretic view. Applied Mathematical Modelling, 38(24), 6098-6107.

Reindorp, M., Fu, M., 2011. Dynamic lead time promising. Proceeding of the 2011 IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning (ADPRL), IEEE, Paris, France, 176-183.

Rohde, J., Meyr, H.,, Wagner, M., 2000. Die supply chain planning matrix. Darmstadt Technical University.

Roitsch, M., Meyr, H., 2015. Oil industry. In: Stadtler, H., Kilger, C., Meyr, H., (Eds.). Supply Chain Management and Advanced Planning. Berlin: Springer, 399-414.

Ruff, V., 2014. Coordinated planning in revenue management. Dissertation, Universität Mannheim.

Samii, A.-B., Pibernik, R., Yadav, P., 2011. An inventory reservation problem with nesting and fill rate-based performance measures. International Journal of Production Economics, 133(1), 393-402.

Samii, A.-B., Pibernik, R., Yadav, P., Vereecke, A., 2012. Reservation and allocation policies for influenza vaccines. European Journal of Operational Research, 222(3), 495-507.

Slotnick, S., 2011. Optimal and heuristic lead-time quotation for an integrated stell mill with a minimum batch size. European Journal of Operational Research, 210(3), 527-536.

Smith, B., Leimkuhler, J., Darrow, R., 1992. Yield management at american airlines. Interfaces, 22(1), 8-31.

Schneeweiß, C., 2003. Distributed Decision Making. Berlin: Springer.

Seitz, A., Ehm, H., Akkerman, R., Osman, S., 2016a. A robust supply chain planning framework for revenue management in the semiconductor industry. Journal of Revenue and Pricing Management, 15(6), 523-533.

Seitz, A., Grunow, M., 2017. Increasing accuracy and robustness of order promises. International Journal of Production Research, 55(3), 656-670.

Seitz, A., Grunow, M., Akkerman, R., 2016b. Data Driven Supply Allocation to Individual Customers Considering Forecast Bias. Available at SSRN: https://ssrn.com/abstract=2813835. Submitted to the International Journal of Production Economics.

Stadtler, H., Kilger, C., Meyr, H., 2015. Supply Chain Management and Advanced Planning, Concepts, Models, Software, and Case Studies. Berlin: Springer.

Stadtler, H., 2012. Master planning – supply network planning. In: Stadtler, H., Fleischmann, B., Grunow, M., Meyr, H., Sürie, C., (Eds.). Advanced Planning in Supply Chains. Berlin: Springer, 109-148.

Stadtler, H., Fleischmann, B., Grunow, M., Meyr, H., Sürie, C., 2012. Advanced Planning in Supply Chains. Berlin: Springer.

Steiner, G., Zhang, R., 2011. Minimizing the weighted number of tardy jobs with due date assignment and capacity-constrained deliveries. Annals of Operations Research, 191(1), 171-181.

Surowiecki, J., 2005. The Wisdom of Crowds. Anchor.

Talluri, K. T., Van Ryzin, G. J., 2004. The Theory and Practice of Revenue Management. New York: Springer.

Talluri, K. T., Van Ryzin, G. J., Karaesmen, I. Z., Vulcano, G. J., 2008. Revenue management: models and methods. Proceedings of the 40th Winter Simulation Conference, Miami, FL.

TATA Consultancy Services, 2013. The Changing Face of Global Order Management. Available at: http://www.tcs.com/resources/white_papers/Pages/Next-Gen-Global-Order-Management.aspx. Accessed: 23.05.2016

Tempelmeier, H., 2008. Material-Logistik – Modelle und Algorithmen für die Produktionsplanung und -steuerung in Advanced Planning Systemen, Berlin: Springer.

Thonemann, U. W., 2002. Improving supply-chain performance by sharing advance demand information. European Journal of Operational Research, 142(1), 81-107.

Tsai, K.-M., 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.

Tsay, A. A., 1999. The quantity flexibility contract and supplier–customer incentives. Management Science, 45(10), 1339–1358.

Tsay, A. A., Nahmias, S., Agrawal, N., 1999. Modeling supply chain contracts: A review. In: Tayur, S., Ganeshan, R., Magazine, M., (Eds.). Quantitative Models for Supply Chain Management. Boston: Springer, 299-336.

Tsay, A. A., Lovejoy, W. S., 1999. Quantity flexibility contracts and supply chain performance. Manufacturing and Service Operations Management, 1(2), 89-111.

Vogel, S., 2014. Demand Fulfillment in Multi-Stage Customer Hierarchies. Wiesbaden: Springer.

Wang, D., Tang, O., 2014. Dynamic inventory rationing with mixed backorders and lost sales. International Journal of Production Economics, 149(1), 56-67.

Wang, D., Tang, O., Huo, J., 2013a. A heuristic for rationing inventory in two demand classes with backlog costs and a service constraint. Computers and Operations Research, 40(12), 2826-2835.

Wang, Y., Zhang, S., Sun, L., 2013b. Anticipated rationing policy for two demand classes under service level constraints. Computers and Industrial Engineering, 65(2), 331-340.

Wight, O. W., 1984. Manufacturing resource planning, MRP II: Unlocking America's productivity potential. New York: John Wiley and Sons.

WSTS Inc., 2015. World Semiconductor Trade Statistics. Available at: https://www.wsts.org/. Accessed: 08.04.2015.

Xiao, T., Shi, J. J., 2016. Pricing and supply priority in a dual-channel supply chain. European Journal of Operational Research (forthcoming).

Yang, Y., 2014. Demand Fulfilment Models for Revenue Management in a Make-to-Stock Production System. Dissertation, Universität Mannheim.

Yang, W., Fung, R., 2014. An available-to-promise decision support system for a multi-site make-to-order production system. International Journal of Production Research, 52(14), 4253-4266.

Yin, Y., Wu, W., Cheng, T., Wu, C., 2014. Due-date assignment and single-machine scheduling with generalised position-dependent deteriorating jobs and deteriorating multi-maintenance activities. International Journal of Production Research, 52(8), 2311-2326.

Zhang, F., Roundy, R., Cakanyildirim, M., Huh, W. T., 2004. Optimal capacity expansion for multi-product, multi-machine manufacturing systems with stochastic demand. IIE Transactions, 36(1), 23-36.

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

Appendix

A Customer segmentation: Minimum segment size

The customer segmentation model presented in Section 7.2.2 sets the minimum segment size s_{min} to $\left\lfloor \frac{2-|I|-|K|}{2(1-|K|)} \right\rfloor$, which maximises s_{min} under the constraint that, if a number of $|K|$ customer segments $k \in K$ of this size are distributed over a sequence S of customers $i \in I$, the segments have a number of at least $(s_{min} - 1)$ unassigned customers in between each other.

The value results, if the s_{min} last customers and the s_{min} first customers of S are assigned to the first two segments and the remaining $|K| - 2$ segments are built by starting with the s_{min}th customer after the last customer that was assigned to the last segment before and assigning s_{min} consecutive customers.

Obviously, s_{min} is maximised by the maximum integer value that fulfils Condition A.1.

$$|I| - 2s_{min} - (|K| - 2)(2s_{min} - 1) \geq 0 \tag{A.1}$$

Solving Condition A.1 and considering the integer condition on s_{min} leads to a value of $\left\lfloor \frac{2-|I|-|K|}{2(1-|K|)} \right\rfloor$.

The value is used in the customer segmentation approach in Section 7.2.2 because it sets a rather large minimum segment size, but, assumed that $|K| \geq 3$, still leaves flexibility to the model to optimise the overall maximum segment width $\bar{w}$.

B Case study cumulated ATP: Material flows and order lead times

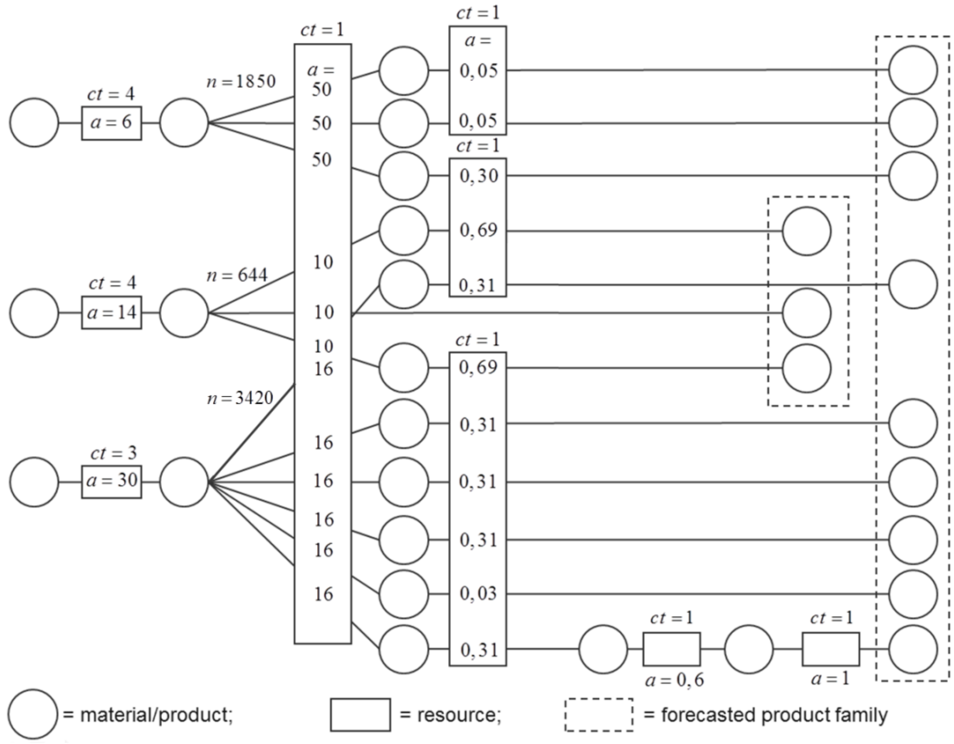

Figure 27: Material flow diagram for product line I

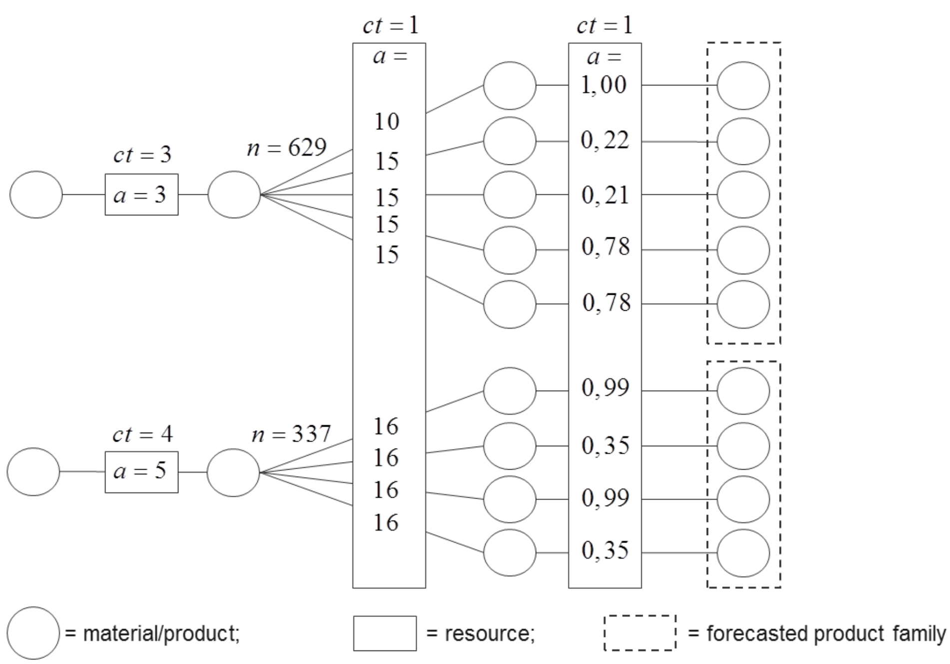

Figure 28: Material flow diagram of product line II

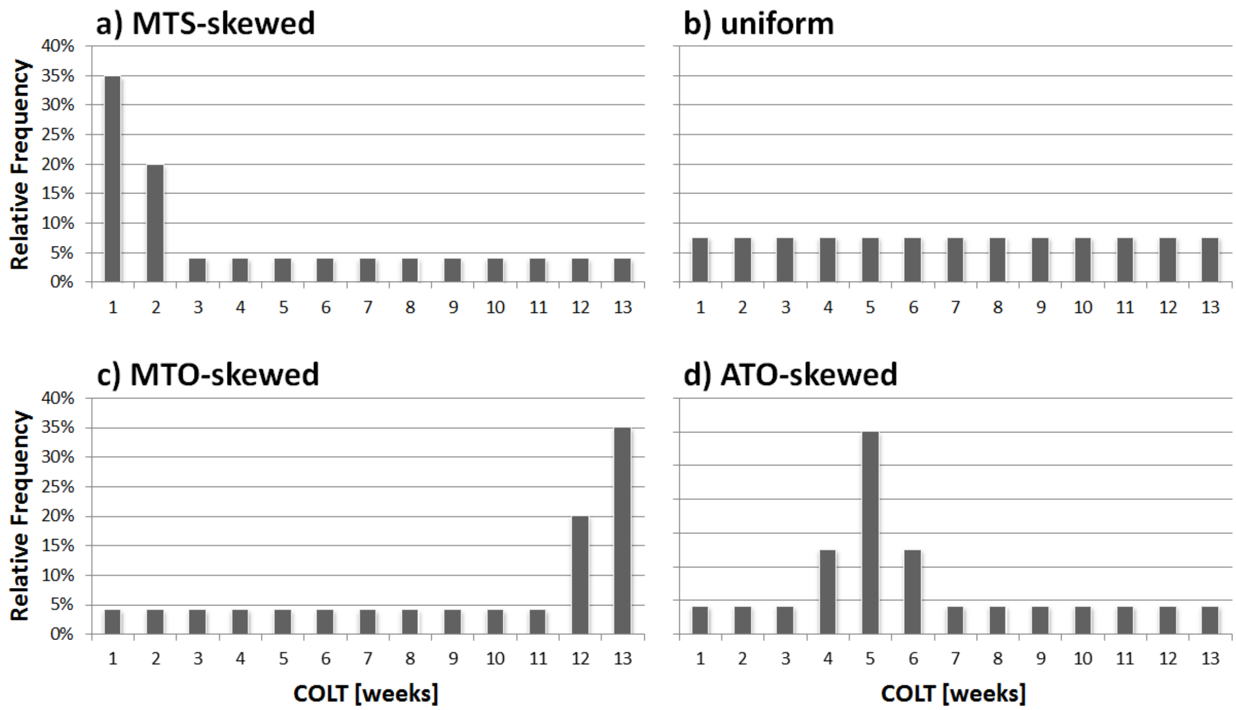

Figure 29: Customer order lead time profiles