Technische Universität München

Lehrstuhl für Logistik und Supply Chain Management

Essays on dynamic production and safety stock planning

Dariush Tavaghof Gigloo

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.

Vorsitzender:		Prof. Dr. Rainer Kolisch
Prüfer der Dissertation:	1.	Prof. Dr. Stefan Minner
	2.	Prof. Dr. Martin Grunow

Die Dissertation wurde am 25.06.2018 bei der Technischen Universität München eingereicht und durch die Fakultät für Wirtschaftswissenschaften am 15.11.2018 angenommen.

Anama həsr edirəm.

To my mother.

Acknowledgements

First and foremost, I owe my deepest gratitude to my supervisor, Prof. Dr. Stefan Minner, for his continuous support, precious guidance and encouragement throughout my PhD. The state-of-the-art knowledge he shared with me is invaluable. His boundless elucidative and new ideas have been very inspiring and helpful to me. His continuous constructive feedbacks and patience paved the way and enabled me to develop this thesis. I am also very grateful for having been given the opportunity to work under his supervision on two industrial research projects from the University of Vienna. I would further like to thank Prof. Dr. Martin Grunow for taking the time to review my dissertation and for being part of the examination committee, as well as to Prof. Dr. Rainer Kolisch for being the chairman of the examination committee.

I would like to express my thanks to my friends and former colleagues at the chair of Logistics and Supply Chain Management at the Technische Universität München: Yuka Akasaka, Szymon Albinski, Dr. Christian Bohner, Dr. Maximilian Budde, Dr. Pirmin Fontaine, Dr. Miray Öner Közen, Sebastian Malicki, Christian Mandl, Thitinan "Kate" Pholsook, Jun.-Prof. Dr. Anna-Lena Sachs, Dr. Florian Taube. I would also like to thank Evelyn Gemkow for the continuous proofreading of my work. I am very thankful to Dr. Lena Silbermayr and Dr. Steffen Klosterhalfen for their collaboration on two papers.

Furthermore, I am very grateful for the support provided to me by my colleagues at Keller & Kalmbach, and I am especially thankful to Dr. Florian Seidl and Dr. Ingomar Schubert for their trust in me during the past two years.

I would like to thank to my close friends, especially to dear Mischan, for supporting me throughout this process. Many thanks to dear Georgina for proofreading my work.

I am deeply grateful to my parents, my sisters and brothers, to whom I definitely owe this work. As far back as I can remember, I have always had their unconditional love, support and encouragement. Without them, I might have never pursued a PhD. I really appreciate the patience they have shown for the fact I have not been able to visit them as much as I would have liked to over the last few years. I would like to thank my lovely niece, Helia, for the patience she showed when she spent six months with me in Munich last year and I wish her all the best in her new academic career.

Abstract

This thesis addresses three essays on production planning under uncertainty.

First, the focus is on a real-world capacity planning problem faced by a highly automated electronics manufacturer, where customer demand is highly volatile. Based on aggregate production planning, a robust and computationally efficient mixed-integer linear program is developed and applied in order to set capacity levels such that a cost-optimal trade-off between flexibility instruments, subcontracting and inventories is achieved.

Second, this thesis considers the stochastic capacitated lot-sizing problem. An optimal solution of this problem requires the integration of dynamic safety stock planning into lot-sizing. Most of the literature and Advanced Planning Systems treat these problems separately and sequentially, even though they are closely interrelated. A new integrated model with service level constraints in the form of a mixed-integer linear program is proposed. The integrated model endogenously sets dynamic safety stocks over non-equidistant lengths of replenishment lead-times. This integrated model is extended to account for re-planning opportunities under rolling horizon planning. An experimental study reveals that the integrated model provides more robust and appealing results than the widely-used sequential approaches because it achieves identical service levels with lower inventories. It is further found that, if there exists sufficient flexibility under rolling horizon planning, the integrated model has to take re-planning opportunities into account in order to avoid the build-up of excess safety stock.

Third, this thesis addresses a real-world stochastic general lot-sizing and scheduling problem from the process industry where stochastic demand is serially correlated. Four mixed-integer linear programs with an increasing degree of sophistication are proposed, to deal with this complex problem. The first two represent a widely-used sequential approach with exogenous safety stocks determined according to a rule-of-thumb or a cost optimisation. The third one is an integrated model that minimises the total costs by simultaneously determining lot-sizes, detailed scheduling and endogenous dynamic safety stock. The fourth variant extends the integrated model to take serially-correlated demand into account. The cost-saving potential of using these approaches is reported in an increasing degree of sophistication based on a real-world dataset under rolling horizon planning. It is found that fast solutions and quick wins in the form of cost-savings of up to 10% can be obtained by using cost-optimised exogenous safety stocks instead of the widely-used rule-of-thumb approach. Increasing the level of model sophistication by using the newly proposed integrated models can lead to further substantial cost-savings of up to 20% and provide more robust results than the widely-used sequential approaches.

Contents

List of Figures **xii**

List of Tables **xiv**

Acronyms **xvi**

1 Introduction **1**

1.1 Motivation . . . 1

1.2 Problem Description . . . 2

1.3 Outline . . . 6

2 Literature Review **9**

2.1 Aggregate Production Planning . . . 9

2.2 Modelling Approaches for the Stochastic Lot-Sizing Problem . . . 11

2.3 Planning Approaches for the Stochastic Lot-Sizing Problem . . . 13

3 Mixed-Integer Linear Programming Formulation for Flexibility Instruments in a Capacity Planning Problem **15**

3.1 Introduction . . . 15

3.2 Problem Description . . . 17

3.3 Model Formulation . . . 19

3.4 Computational Results . . . 24

3.4.1 Case Study . . . 24

3.4.2 Results for Randomly Generated Instances . . . 27

3.5 Conclusions . . . 31

4 *Planning Approaches for Stochastic Capacitated Lot-Sizing with Service Level Constraints* **36**

4.1 Introduction 37
4.2 Problem Description 40
4.3 Stochastic Dynamic Program 42
4.4 Mixed-Integer Linear Programs 45
4.4.1 Sequential Approach 45
4.4.2 Integrated Model 46
4.4.3 The Integrated Model with Re-planning-Opportunity Adjustment . 57
4.5 Numerical Study 61
4.5.1 Experimental Design 61
4.5.2 Numerical Results 65
4.6 Conclusions 74

5 Integrated Lot-sizing, Scheduling and Dynamic Safety Stock Planning: A Real-World Case Study **76**

5.1 Introduction 77
5.2 The Production Setting and Current Planning Practice of the Company under Study 80
5.3 General Problem Description, Assumptions and Notation 83
5.4 Calculation of Forecast Bias and Demand Uncertainty 84
5.5 Lot-sizing and Scheduling with Exogenous Safety Stock Targets 86
5.5.1 Rule-of-Thumb for Safety Stock Targets 86
5.5.2 Cost-based Approach for Safety Stock Targets 86
5.5.3 Mixed-Integer Linear Program for Lot-sizing and Scheduling with Exogenous Safety Stocks 88
5.6 Integrated Lot-sizing, Scheduling and Dynamic Safety Stock Planning Model 92
5.6.1 Shortage-Cost Minimisation Approach 92
5.6.2 Demand Uncertainty Parameters over Replenishment Cycles 94
5.6.3 Re-planning-Opportunity Adjustment under Rolling Horizon Planning 100
5.6.4 Linearisation of Expected Backlogged Quantity 105

5.7 Solution Approaches . . . 112
5.7.1 Period-based Decomposition Heuristic . . . 112
5.8 Numerical Analysis . . . 113
5.8.1 Data . . . 113
5.8.2 Evaluation Approach . . . 114
5.8.3 Results . . . 115
5.9 Conclusions . . . 122

6 Conclusions **124**
6.1 Summary . . . 124
6.2 Limitations and Future Research . . . 127

Bibliography **129**

List of Figures

1.1 Interrelation of safety stocks and lot-sizes. 4

3.1 Multi-item multi-facility capacity planning problem with parallel machines. 18

4.1 Approximation grid: Selecting a lower triangle (left-hand side), selecting an upper triangle (right-hand side). Indexes p and t have been dropped. . 54

4.2 Replenishment cycle of product p with an average length of L_p. 58

4.3 Rolling horizon setting. 61

4.4 Evolution of the average inventory level ($\beta^c = 0.90$). 67

4.5 The integrated model and the re-planning opportunity adjusted integrated model versus the sequential approach: The development of the average of ΔTC in %. 71

4.6 Choosing an appropriate modelling approach with respect to the planning approach and capacity level. 75

5.1 Replenishment cycle. 78

5.2 Historical demand time series for the different products. 82

5.3 Correlation in the Variance of the Replenishment Cycle Demand (VRCD). 98

5.4 Illustration of the determination of $\mathbf{m}^{\mathrm{crl}}_{pt}$. 99

5.5 The scaling factor $\boldsymbol{\alpha}_{p,t,\tau}$ in period $t = 2$. 101

5.6 Approximation grid to linearise the expected backlog function based on the triangulation method. 108

5.7 The improved approximation grid. 109

5.8 Period-based decomposition heuristic. 113

5.9 Cost-saving potential due to re-planning opportunity adjustment. 122

5.10 Cost-saving potential of different planning approaches compared to the rule-of-thumb. 123

List of Tables

3.1 Notation. 20
3.2 Flexibility instruments in cases. 25
3.3 Total production cost improvement (in %) due to flexibility instruments of the case study. 27
3.4 Computational performance of the proposed models of case study. 28
3.5 Overview of sets and parameters of case study and randomly generated instances. 30
3.6 Total production cost improvement. 32
3.7 Summary of computational performance: Instances solved to optimality, avg. runtime of instances solved to optimality and avg. optimality gap of instances not solved to optimality within 1 hour. 33
3.8 Runtime of Cbc relative to Xpress. 34
3.8 Runtime of Cbc relative to Xpress. 35

4.1 Notation. 42
4.2 Notation for the sequential approach. 45
4.3 New decision variables for the integrated model. 51
4.4 New notation for demand uncertainty parameters. 51
4.5 Mean and variance of replenishment cycle demand. 53
4.6 Notation used for the approximation grid. 56
4.7 Experimental design. 62
4.8 Computational performance (runtime in seconds, gap in %). 66
4.9 Soft service level constraints versus hard service level constraints. 66
4.10 Realised service level (β^{RL}) in %. 68

4.11 The integrated model versus the sequential approach (ΔTC in %). 70
4.12 The re-planning opportunity adjusted integrated model versus the sequential approach for the uncapacitated case under rolling horizon planning (ΔTC in %). 71
4.13 Integrated model versus stochastic dynamic program under rolling horizon planning. 74

5.1 Notation for the sequential approach. 91
5.2 Additional notation for the integrated approach*. 95
5.3 Notation used for the approximation grid. 111
5.4 Comparison of Full Mixed-Integer Linear Program (F-MILP) with period-based Decomposition Heuristic (DH). 117
5.5 Cost overview of the different planning approaches. 119
5.6 Replenishment cycle length comparison for first replenishment cycle in 2014.121

Acronyms

APP Aggregate Production Planning

APS Advanced Planning Systems

cDoS common Days-of-Supply

CLSP Capacitated Lot-Sizing Problem

DH period-based Decomposition Heuristic

DoS Days-of-Supply

EMS Electronics Manufacturing Services

GAG General Approximation Grid

GD Gamma Distribution

GLSP General Lot-sizing and Scheduling Problem

F-MILP Full Mixed-Integer Linear Program

HSLC Hard Service Level Constraint

IAG Improved Approximation Grid

iDoS individual Days-of-Supply

IM Integrated Model

IM-CRRL Integrated Model with serial demand CoRReLation

LP Linear Program

MILP Mixed-Integer Linear Program

MIP Mixed-Integer Programming

MRP II Manufacturing Resource Planning

MRCD Mean of the Replenishment Cycle Demand

ND Normal Distribution

RoT Rule-of-Thumb

SA Sequential Approach

SDP Stochastic Dynamic Program

S-CLSP Stochastic Capacitated Lot-Sizing Problem

S-GLSP Stochastic General Lot-sizing and Scheduling Problem

SSLC Soft Service Level Constraint

S-ULSP Stochastic Uncapacitated Lot-Sizing Problem

TC Total Cost

VRCD Variance of the Replenishment Cycle Demand

WIP Work In Process

Chapter 1

Introduction

1.1 Motivation

Global sourcing, exploiting economies of scale and uncertainty in demand and supply force many companies to build up inventories for raw materials, as well as semi-finished and finished products throughout the supply chain. According to the 27th Annual State of Logistics Report©, the value of the inventories of firms in the USA was more than $2.5 trillion in 2015 (AT Kearney, 2016). The management of these tremendous inventories opens up new opportunities and challenges when it comes to achieving and maintaining competitive advantages in a highly uncertain business environment. The right decisions must be made on the quantities, timing and location of the production and replenishment throughout the supply chain. On the one hand, this is necessary in order to provide a high customer service level, but on the other hand, the companies want to avoid unnecessary and obsolete inventories.

A main characteristic of these environments is uncertainty, which is either caused by the customer demand or by unreliable production or supply (Minner, 2000). Uncertainties have increased as a result of the diversification of products and customer expectations, shortening product life cycles and longer replenishment cycles due to global sourcing (Silver et al., 2017).

The two most common approaches of addressing uncertainty are the establishment of flexibility and the deployment of inventories as safety stocks. A flexible supply chain can be obtained by incorporating information and communication technologies, new manufacturing technologies, capacity adjustments and contracting mechanisms (Bertrand, 2003). While economies of scale and global sourcing can limit the degree of flexibility, safety stocks become a vital element for maintaining the customer service level. From a hierarchical planning perspective (Schneeweiss, 2003), decisions regarding the type and degree of flexibility are usually made on a strategic level. At the lower levels, i.e., in the medium or short-terms, the question is how to optimally utilise the existing flexibility instruments. At these levels, flexibility becomes fixed and limited, which means that the use of safety stocks becomes more important for tackling uncertain demands in order to maintain the desired customer service level.

There are some major pitfalls when it comes to the implementation of flexibility and safety stocks. The impact of flexibility, in terms of capacity levels on safety stocks, is usually ignored (Mapes, 1993). Moreover, safety stocks are traditionally decoupled from decisions on quantities and the timing of production and replenishment, even though they are closely interrelated with each other. In industrial practice, these simplified approaches provide fast and easily-implementable solutions at the expense of sub-optimality in terms of the utilisation of vital resources and inventories. Thus, overcoming those shortcomings can substantially improve the customer service level and avoid unnecessary inventories.

To this end, this thesis focuses on (i) the impact of using different flexibility options on the adjustment of capacity levels with respect to customer demand fluctuations, (ii) simultaneous decision-making on safety stocks, the quantities and the timing of production and replenishment and (iii) the impact of capacity levels on the safety stock requirement.

1.2 Problem Description

Recent developments in information and communication technology, along with products with shortened life cycles and mass-customisation have made flexibility one of the most important performance indicators in supply chains. The exploitation of the existing flexibility results in the better and faster adjustment of internal resources when faced with external fluctuations and changes.

The first problem addressed in this thesis relates to Aggregate Production Planning (APP), which analyses the acquisition and allocation of internal resources with limited capacities in order to fulfil customer demand in a finite planning horizon (Graves, 2011). The planning horizon is usually medium to short-term and divided into discrete periods with aggregated customer demands and internal resource capacities. APP usually takes the workforce level, overtime scheduling, subcontracting and the inventory level into account and looks for the best trade-off between these factors in response to the customers' varying demand.

APP has received a lot of attention in the literature (see Chapter 2 for related literature). Our contribution is a practical application in the real-world case study of a highly automated electronics manufacturer. Flexibility instruments, such as shift planning, overtime account and maintenance planning, are incorporated separately or in combination with each other. The robustness of the proposed modelling approach is important in terms of the quality of the solutions and the required computational effort.

From a hierarchical planning point of view, APP provides capacity levels for key internal resources as inputs for generating more disaggregated plans. From a medium to short-term perspective, the disaggregated plans provide the quantities and timing of production and replenishment, which are the solution to the lot-sizing and scheduling problems. The lot-sizing and scheduling problems are observed in a wide variety of disciplines and industrial sectors, including consumer products, electronics, pharmaceuticals, steel, and the chemical industry. The approaches used for tackling them are vital components of planning software, such as Manufacturing Resource Planning (MRP II) and Advanced Planning Systems (APS).

The second problem addressed in this thesis focuses on a single-stage big bucket Stochastic Capacitated Lot-Sizing Problem (S-CLSP). In this problem, (final) products of an uncertain demand are produced directly from raw materials (without sub-assemblies) on a single machine with limited capacity. The aim is to find cost-optimal decisions regarding when and how much of each product should be produced by taking the limited capacity of the machine into account in order to satisfy uncertain demand with respect to a prescribed target service-level. In S-CLSPs, the degree of flexibility in terms of capacity levels is usually fixed and limited. In these problems, holding safety stocks is a vital tool for protecting against uncertain demand.

The S-CLSPs are usually solved sequentially (see, e.g., de Kok and Fransoo, 2003). In planning software, such as MRP II and APS, it is common to first solve a safety stock planning problem in order to capture the demand uncertainty and to then solve a deterministic lot-sizing (scheduling) problem that is decoupled from uncertainty but takes safety stocks into account as a constraint. The main pitfall of such a sequential approach is that it does not take the interrelation of the decisions on safety stocks or those on quantities and timing of production and replenishment into account.

Safety stocks are placed to buffer against the demand uncertainty over *replenishment cycles*. The replenishment cycle depends on the production plan, which is the timespan between two consecutive production periods. As an example, the production plan for one product over twelve time periods illustrated in Figure 1.1 is discussed in the following. Production occurs in periods 1 and 6, marked in red. In the remaining periods, marked in blue, this product is not produced. Therefore, the first replenishment cycle extends over 5 periods. It starts at the beginning of period 1 and ends at the end of period 5. Thus, safety stock planned in period 1 must cover demand uncertainty over the 5 periods. If the next lot were to be produced as early as in period 4, instead of in period 6, the replenishment cycle would only be 3 periods. Thus, the production plan clearly impacts the safety stock requirement. On the other hand, the safety stock quantities need to be produced at some point. Therefore, the safety stock requirement also affects the production planning problem. Both are interrelated.

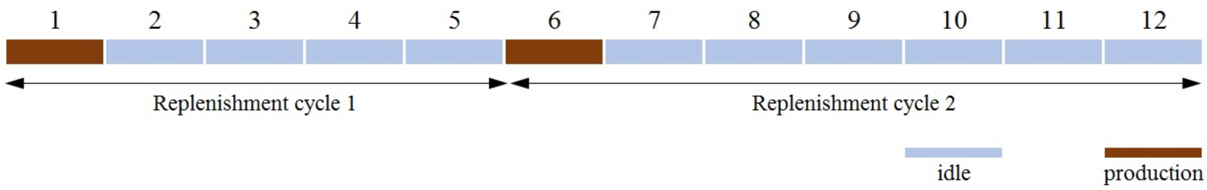


Figure 1.1: Interrelation of safety stocks and lot-sizes.

Despite this major shortcoming, the main advantage of the sequential approaches is that they reduce the complexity of using an integrated model that addresses both problems simultaneously. Schneeweiss (2003) refers to this type of decomposition as a constructional distributed decision making system where a complex operations research problem is divided into more manageable sub-problems. Using separate approaches to determine safety stocks (a higher level) and to produce plans (a lower level) reduces the complexity of the original problem.

A simultaneous consideration of both problems has recently attracted increasing attention. The existing literature makes considerable contributions to the S-CLSPs (see Chapter 2 for related literature). However, the main aspects of these problems are rarely discussed concurrently: The existing literature either neglects capacitated cases, assumes exogenous replenishment cycles, or focuses on production times and disregards production quantities. From a modelling perspective, an endogenous determination of replenishment cycles has not been fully solved. The existing literature mainly uses a complete enumeration of every possible length of replenishment cycles, which restricts them to small-size and, particularly, to discrete problems.

Moreover, in practice, the production plans are predominantly implemented under rolling horizon planning. This means that, based on a demand forecast, first, a finite planning horizon problem is solved. The production plan of some initial periods is implemented. The planning horizon is rolled forward and, at the same time, demand forecast is updated as more information is gained over time and the system status is updated as the actual demand is realised. The main advantage of a rolling horizon planning approach is that it provides re-planning opportunities for the solution during periods that are not immediately implemented. In other words, based on the realised demand and updated demand forecast, there are opportunities to adjust the original plan (re-plan) over the course of the following (rolled) planning. Another shortcoming of the integrated models found in the literature is that they ignore the re-planning opportunities which exist under rolling horizon planning. This calls for a careful and fair investigation on whether using such an integrated model instead of a sequential approach can really pay off under rolling horizon planning. After all, solving integrated approaches requires more computational efforts than solving the sequential approaches.

Nevertheless, an optimal solution for stochastic lot-sizing problems under rolling horizon planning can be obtained by using a Stochastic Dynamic Program (SDP) (Mula et al., 2006). However, the SDPs suffer from the curse of dimensionality. Consequently, this method is of only limited use in practice since the industrial lot-sizing (scheduling) models usually have large problem sizes and various production constraints with real-world settings (Jans and Degraeve, 2008). For solving these types of problems, Mixed-Integer Linear Programs (MILPs) are usually applied. For practical purposes, it is desirable to find a way to consider the re-planning flexibility in an MILP.

The final problem addressed in this thesis is a real-world Stochastic General Lot-sizing and Scheduling Problem (S-GLSP) from the process industry. This problem refers to the stochastic version of a simultaneous big-bucket and small-bucket lot-sizing and scheduling problem where customer demand is uncertain and serially correlated over a planning horizon. The deterministic counterpart of this problem, known as the General Lot-sizing and Scheduling Problem (GLSP) (Fleischmann and Meyr, 1997), integrates short-term detailed scheduling into lot-sizing. This is the most general setting and the other lot-sizing and/or scheduling problems are usually seen as special cases of this version. In the stochastic version of this problem, dynamic safety stocks, lot-sizing and scheduling are determined simultaneously. Specifically, lot-sizes are determined on a macro level with discrete and equidistant periods, scheduling is done on a micro level and based on continuous and non-equidistant periods and dynamic safety stocks are endogenously placed on a micro level over continuous non-integer replenishment cycles. Despite its high practical relevance, the S-GLSP has not yet been addressed in the literature.

This thesis contributes to the huge body of existing literature on the stochastic lot-sizing (and scheduling) problems by answering the following main research questions:

(i) Using an MILP, how can dynamic safety stocks be endogenously determined over non-equidistant (continuous) lengths of replenishment cycles in the S-CLSP and the S-GLSP?

(ii) What is the impact of the capacity level on the performance of sequential and integrated modelling approaches under rolling horizon planning?

(iii) Using an MILP, how can serially-correlated demand be taken into account in order to endogenously adjust the dynamic safety stock levels and what is the implication of doing so?

(iv) What are the cost-saving potentials if practitioners increase the planning maturity level by using more sophisticated modelling approaches to address a complex real-world S-GLSP?

1.3 Outline

The remainder of this thesis is structured as follows. Chapter 2 presents the related literature on the APP and the stochastic lot-sizing problem. Chapter 3 explores the flexibility instruments used for adjusting capacity levels in order for them to correspond with demand fluctuations in a real-world case study of a highly automated electronics manufacturer. This problem assumes a multi-item, multi-facility, multi-stage capacity planning problem with parallel machines. Three flexibility instruments are introduced: Shift planning, overtime account and flexible maintenance. In the context of APP, an MILP is developed and applied. It sets capacity levels while a cost optimal trade-off between flexibility instruments and inventories is obtained. An extensive numerical study based on a real-world dataset is conducted to evaluate the robustness of the modelling approach, as well as the cost-saving potential of the available flexibility instruments. The computational performance of the modelling approach is validated by two commercial and non-commercial MILP solvers. The numerical study reveals that the most substantial cost-saving can be obtained through shift planning. The combination of all these flexibility instruments is crucial to manufacturers who face high demand variabilities, especially when production capacity is limited. This chapter is based on Tavaghof-Gigloo et al. (2016).

Chapter 4 addresses the single-stage big-bucket Stochastic Capacitated Lot-Sizing Problem (S-CLSP). The planning horizon is finite with equidistant discrete time periods. Customer demand is a random, non-stationary and an independent variable during the planning horizon, with a known probability distribution and a given mean and variance per product and per period. Customer demand is backlogged if it is not met by the end of a period. There is a linear inventory holding cost for keeping one unit of a product at the warehouse. The production capacity of the machine is limited. To produce a product, the machine must be set up with the required setup time. This problem is described by using an SDP. Then, three MILPs are presented to address this problem: First, a sequential approach that assumes a deterministic capacitated lot-sizing problem with predetermined safety stocks is studied. Secondly, an integrated model where dynamic safety stocks and lot-sizes are simultaneously determined is introduced. This approach is based on a chance-constrained program and incorporates a fill-rate service level. If capacity is limited, soft service level constraints are introduced to avoid infeasibility.

Thirdly, an integrated model that takes re-planning opportunities, which exist in a rolling horizon approach, into account is proposed. In an experimental study, various main input factors are defined in order to compare the cost-saving potentials of the modelling approaches in respect to the realised identical customer service level. The experimental study reveals that using the integrated model instead of the sequential approach can only be justified if flexibility is strictly limited under rolling horizon planning. If sufficient safety capacities exist, using an integrated model only pays off if re-planning opportunities are further taken into account in a rolling horizon approach. This chapter is based on Tavaghof-Gigloo and Minner (2019).

Chapter 5 studies a real-world Stochastic General Lot-sizing and Scheduling Problem (S-GLSP) from the process industry. The main assumptions and constraints are the existence of sequence-dependent setup times and costs, minimum production quantities and times, inventory and production capacity constraints and back-ordering of unfulfilled demand, where demand is further serially-correlated over the planning horizon. With an increasing degree of sophistication, four MILPs are introduced in order to deal with this complex problem; (i) a sequential approach with simple rule-of-thumb exogenous safety stocks, (ii) a sequential approach with cost-optimised exogenous safety stocks, (iii) an integrated model based on a cost-minimisation method and (iv) an integrated model that further takes the serially-correlated demand into account. Since even the deterministic counterpart of this problem is hard to solve, a period-based decomposition approach is introduced in order to obtain promising results in a reasonable amount of time. Based on a real-world dataset and sensitivity analyses, the cost-saving potentials of the integrated approaches are compared to less sophisticated sequential approaches. The results show that, in the sequential approaches, fast solutions and quick cost-savings of up to 10% can be achieved by replacing the widely-used rule-of-thumb approach for the exogenous safety stocks with the cost-optimised ones. Increasing sophistication in terms of planning by our industrial partner in order to use the proposed integrated approaches can unlock further substantial cost-saving potentials of up to 20% and provide more robust results than the widely-used sequential approaches. This chapter is based on Tavaghof-Gigloo et al. (2019).

Chapter 2

Literature Review

This chapter is structured as follows. Section 2.1 presents the related literature on Aggregate Production Planning (APP). Section 2.2 and 2.3 present prior works on stochastic lot-sizing problems. In Section 2.2, we focus on the mathematical modelling approaches proposed in the literature for stochastic lot-sizing problems. In Section 2.3, we review the recent literature on the planning approaches for stochastic lot-sizing problems.

2.1 Aggregate Production Planning

An early extensive literature review on the APP has been given by Nam and Logendran (1992). Silva et al. (2000) study an APP model where the workforce level can be adjusted at the beginning of the planning horizon and remains unchanged afterwards. Lagodimos and Mihiotis (2006) study shift planning with overtime accounts. Their results show that an effective use of overtime leads to workforce reductions and improved utilisation. Da Silva et al. (2006) develop a multi-criteria Mixed-Integer Linear Program (MILP), i.e., maximising profit, minimising late orders, minimising changes of the workforce level, by taking constraints of production, inventory and workforce into account. They embed the developed model into a decision support system for practical usage.

Othman et al. (2012) present a multi-objective non-linear programming model to determine the workforce level and overtime hours. Ramezanian et al. (2012) introduce an MILP for a two-stage APP problem and apply a genetic algorithm and tabu search to solve the problem. Askar and Zimmermann (2007) and Askar et al. (2007) address a capacity adaptation and staff planning problem on a single assembly line in the automotive industry by taking cycle time, shift planning, work regulations and line balancing into account. They propose a solution approach based on dynamic programming and show that some real world problems can be solved efficiently. Sillekens (2008) and Sillekens et al. (2011) introduce an MILP of the problem studied by Askar and Zimmermann (2007) and extend it to include buffers between shops. Walter et al. (2011) evaluate volume flexibility instruments by using a three-step method that consists of a preliminary analysis where the effective flexibility instruments are identified. They also introduce an optimisation model and design-of-experiments techniques for evaluating the flexibility instruments. Hemig et al. (2014) discuss an integrated production and staff planning problem for heterogeneous, parallel assembly lines in the automotive industry, also proposing a dynamic programming solution approach. Merzifonluoğlu et al. (2007) develop a class of production planning models that integrates subcontracting and overtime options. They provide effective solution methods by using polyhedral properties and dynamic programming techniques.

Another area of research relevant to this topic is the integration of maintenance planning into the APP. Typically, production planning and maintenance planning are considered separately. While production planning models try to balance the total costs of production and inventory, maintenance planning models usually aim at balancing the costs and benefits of maintenance in order to ensure a reliable production system. Occasionally, the two goals contradict each other (Aghezzaf et al., 2007). Hence, only a few papers address the integration of maintenance and APP. Weinstein and Chung (1999) propose a three-stage model for evaluating the maintenance policies of an organisation by integrating the model into APP. First, they generate an aggregate production plan. Then, a master production schedule that minimises the deviations from the specified first-stage aggregate production goals is developed. Finally, they use work centre loading requirements to simulate equipment failures during the planning horizon.

Aghezzaf et al. (2007) discuss a joint production and maintenance planning model for a production system, including random failures for finding an integrated lot-sizing and preventive maintenance strategy. Najid et al. (2011) consider a joint production and maintenance problem and present an MILP that takes demand shortages and reliability of the production line into account. Their model can be solved to optimality for small problems. Allaoui et al. (2011) propose two-level planning to hierarchically integrate production and maintenance planning. They integrate preventive maintenance into the APP and corrective maintenance into the detailed planning. Alaoui-Selsouli et al. (2012) incorporate a Lagrangian relaxation heuristic for solving a joint production planning and maintenance problem. Arts and Flapper (2015) introduce an MILP for an aggregate planning of rotable overhaul and supply chain operations to determine aggregate workforce levels, turn-around-stock levels of modules and overhaul and replacement quantities per period.

2.2 Modelling Approaches for the Stochastic Lot-Sizing Problem

There is a vast body of literature on the modelling and solution approaches of the deterministic lot-sizing problems. Extensive reviews are given by, e.g., Karimi et al. (2003), Jans and Degraeve (2008), Jans and Degraeve (2007) and Copil et al. (2017). In this study, only the major related works that consider stochastic demand are presented.

The literature on the stochastic lot-sizing problems can be classified into the big-bucket and small-bucket stochastic lot-sizing problem. In a big-bucket problem, the length of production periods is long enough to produce multiple products in one period. In a small-bucket problem, only one product is produced in every period.

Taking the big-bucket stochastic lot-sizing problems into account, Silver (1978) introduces a chance-constrained program to address a single-product Stochastic Uncapacitated Lot-Sizing Problem (S-ULSP). He proposes a three-stage heuristic solution approach that sequentially determines reorder points, production cycles and lot-sizes. Askin (1981) uses a two-stage heuristic approach with a simultaneous determination of production cycles and lot-sizes for the same problem. Bookbinder and Tan (1988) introduce three main uncertainty strategies for lot-sizing problems in the presence of uncertain demand: *Static-dynamic uncertainty*, *static uncertainty* and *dynamic uncertainty*.

Under static-dynamic uncertainty, only the production times are determined by a specification of the target inventory levels in the production periods. Accordingly, production quantities are determined as the actual demand is observed over time. Based on this strategy, Tarim and Kingsman (2004) introduce an MILP for the S-ULSP that incorporates the simultaneous determination of the production periods and the order-up-to-levels under α-service level constraints. Tempelmeier (2007) extends this model by including negative inventory levels. Moreover, Rossi et al. (2015) propose the use of MILP models to determine the order-up-to-levels for the single-product S-ULSP under various service levels and a shortage-cost model by introducing a linearisation technique that determines the upper and lower bounds of the first-order loss function.

Under the static uncertainty strategy of Bookbinder and Tan (1988), the production times and quantities for the entire planning horizon are determined at the beginning of the planning horizon. Based on this strategy, Tempelmeier and Herpers (2011) address a single-product S-ULSP under a cycle fill-rate (β^c-service level). They present a solution that is based on a modified shortest-path problem. Tempelmeier (2011) looks into a stochastic version of a multi-product S-CLSP and proposes a heuristic solution approach that combines column generation with the ABC_{β^c} heuristic introduced by Tempelmeier and Herpers (2010). Helber et al. (2013) introduce a multi-product S-CLSP with a new backorder-oriented service level (δ service level) with respect to the stockout duration. Tempelmeier and Hilger (2015) modify this model by taking the cycle β^c-service level into account. They approximate the non-linear functions of the expected on-hand inventory and the expected back-orders by using piece-wise linear functions. They solve the problem with an adjusted fix-and-optimize heuristic.

Finally, under the dynamic uncertainty strategy of Bookbinder and Tan (1988), both the production time and production quantities are determined in every period as demand is realised. Bookbinder and Tan (1988) argue that this so-called *wait and see* strategy is not usually desired in practice and can cause poor cost performance if the setup cost is considerably higher than the inventory holding costs.

Other relevant contributions consider the modelling of production planning problems with limited capacity and uncertain demand. In this stream, mainly non-linear interdependency between resource utilisation, production lead-time and safety stocks is investigated. Orcun et al. (2009) present a chance-constrained program that can be used for integrating production release planning and safety stocks by taking load-dependent production lead-times with clearing functions into account. Ravindran et al. (2011) look into a similar problem and introduce a tractable deterministic heuristic to solve the original stochastic problem. While the model of Ravindran et al. (2011) treats the replenishment cycle as an exogenous parameter, the use of clearing functions in conjunction with the shortfall will yield a dynamic model. Albey et al. (2015) and Albey et al. (2016) integrate the evolution of the forecasted demand into the production planning for stochastic demand. They use a new version of the chance-constrained program that is based on the shortfall approach of Glasserman (1997). As a result of the shortfall approach, they present soft constraints for safety stocks.

A review on small-bucket stochastic lot-sizing, known as the stochastic economic lot-scheduling problem, is given by Winands et al. (2011). Recent approaches for the parameter optimisation in stochastic stationary environments are presented in Löhndorf and Minner (2013) and Löhndorf et al. (2014).

The stochastic version of the simultaneous lot-sizing and scheduling problem that is the focus of Chapter 5 has not yet been addressed in the literature. Its deterministic counterpart, known as the GLSP has been investigated by Fleischmann and Meyr (1997), Koçlar (2005) and Koçlar and Süral (2005). Floudas and Lin (2004) and Transchel et al. (2011) studied the GLSP with continuous and non-equidistant micro periods. Copil et al. (2017) provide a recent review on the simultaneous lot-sizing and scheduling problem.

The serially-correlated demand has not yet been addressed in the literature on common stochastic lot-sizing problems. In the stochastic-demand inventory literature, the serially-correlated demand case has been investigated by authors such as Ray (1980), Charnes et al. (1995), Urban (2000) and Disney et al. (2015).

2.3 Planning Approaches for the Stochastic Lot-Sizing Problem

Despite to the considerable contribution with respect to the modelling approaches for stochastic lot-sizing, the literature on the evaluation of stochastic lot-sizing under rolling horizon planning is limited. Bookbinder and H'ng (1986) and Bookbinder and Tan (1988) investigate a single-product S-ULSP with non-stockout probability in a rolling horizon approach. Bookbinder and Tan (1988) compare the performance of their approach with that of Silver (1978). They observe that their approach leads to a lower total cost than the approach of Silver (1978), which, however, often returns a better realised service level. They further observe an over-achievement of the target service level under rolling horizon planning. Meistering and Stadtler (2017) evaluate an S-CLSP. They first introduce an approach that depends on the time-between-orders in order to place safety stocks. They further introduce a rolling horizon setting to fix the cycle β^c-service level at a given level over an evaluation period by adjusting the target service level in each planning period. They use fairly moderate capacity utilisation cases. Additional external capacity is still required to avoid infeasibility. Moreover, they compare the performance of the different rolling horizon settings with respect to the total cost. They find that their rolling horizon setting can often provide the lowest total cost. However, similar to the observation of Bookbinder and Tan (1988), Meistering and Stadtler (2017) also find that the higher total cost of an approach usually corresponds to a higher realised service level.

Lin and Uzsoy (2016) present a chance-constrained program for a single-product single-stage production planning problem under the assumption that the length of the replenishment cycles is an exogenous parameter. They introduce two different settings for the chance-constraints to update the length of the exogenous replenishment cycle over time. They compare the performance of their settings to a linear programming formulation by Sridharan and Berry (1990) that does not take uncertain demand into account. They find that their settings can reduce schedule instability while maintaining high service levels. Albey et al. (2015) and Albey et al. (2016) develop chance-constrained programs for a production planning problem with demand forecast evolution under rolling horizon planning. One of their main findings is that the advantage of using a more accurate forecasting technique, such as forecast evolution, is more pronounced if there is excess capacity in the system.

Chapter 3

Mixed-Integer Linear Programming Formulation for Flexibility Instruments in a Capacity Planning Problem

In this chapter, we present an MILP for the Aggregate Production Planning (APP) problem of an electronics manufacturer. A multi-item, multi-facility, multi-stage capacity planning problem over a finite planning horizon with deterministic demand is considered. We include the flexibility instruments shift planning, overtime account and flexible maintenance. We present an extensive computational study where the proposed model is applied in a real-world case study and for randomly generated instances. Using a full factorial experimental design we evaluate the cost-saving potential of the flexibility instruments and their combinations. The computational performance of the proposed model formulation is investigated by applying different Mixed-Integer Programming (MIP) solvers.

3.1 Introduction

Nowadays, manufacturing companies have to cope with an increasing demand variation due to shortening product life cycles and seasonality. Demand variation stipulates increasing flexibility in the production environment. APP provides this flexibility by adapting the internal resources to the demand variations. APP determines optimal levels of production, inventory and workforce over a given finite planning horizon by taking restrictions on the demand fulfilment and production resources into account. Therefore, effective and flexible production planning can be essential for the success of a manufacturer in highly competitive manufacturing environments.

This topic was originally motivated by a highly automated electronics manufacturer who produces multiple items at several facilities with different costs in multi-stage production processes with parallel machines and capacity limitations over multiple periods with time-varying future demand forecasts. The manufacturer faces significant demand variations over the planning horizon. The main challenge is to optimise and balance resource utilisations on different production lines. To do so, we introduce a mathematical model formulation that integrates three flexibility instruments into the APP: Flexible shift planning, overtime account and flexible maintenance.

Flexible shift planning and overtime account are two important flexibility strategies in the APP that help manufacturers to adapt supply capacities. Shift planning includes several shift models, each associated with a predetermined shift cost calculated mainly through the required workforce level. Hence, the manufacturer may benefit from flexible shift planning by adapting the workforce level to demand variations. Switching between shift models, however, requires an adaptation of the workforce level, which results in extra cost. In this work, we do not explicitly consider detailed staff planning. Instead, we use flexible shift planning to adapt working time and consequently the workforce level. This aggregation of workforce planning into flexible shift planning makes sense in highly automated manufacturing environments – like the company under study in this chapter – where production characteristics such as the cycle times are not heavily influenced by the workforce level.

In contrast to the typical working time account, which accumulates the differences between the actual working hours and the working hours in the contracts, we introduce an overtime account that captures the interrelation between regular working time and overtime. This flexibility enables the manufacturer to compensate under-time hours with overtime hours and to minimise the total overtime cost with respect to the regulations and collective agreements.

Additionally, integrating flexible maintenance into the APP may also be beneficial for the manufacturer, since it allows for optimal scheduling of interval-based maintenance activities. This integration also minimises the interruption of the production activities due to maintenance when production resources are highly utilised in peak times.

Our contribution is a practical application of a capacity planning model based on a real-world problem in the highly automated electronics industry. To gain managerial insights into the value of the different flexibility instruments in the APP, we implement a full factorial experimental design by defining input factors like the available production capacity, demand variation, shift cost and shift change cost. For practical purposes, we analyse the computational performance of the proposed model, using both commercial and non-commercial solvers. Specifically, we address the following research questions: (1) What is the cost benefit of applying the flexibility instruments in a highly automated manufacturing environment separately or in combination with each other? (2) How robust is the proposed mathematical model formulation in terms of solving the underlying problem under different input factor levels using different MIP solvers?

The remainder of this chapter is structured as follows. Section 3.2 describes the problem and the flexibility instruments. Section 3.3 provides the mathematical formulation of the problem. Section 3.4 presents an extensive numerical study. Finally, Section 3.5 summarises the findings of this chapter.

3.2 Problem Description

We study a multi-item, multi-facility, multi-stage capacity planning problem with parallel machines. Each item has a dynamic deterministic demand over a finite planning horizon divided into discrete time periods. Demand must be fulfilled without any delay, i.e., backlogging is not permitted. If, due to production capacity restrictions, demand cannot be fulfilled, it is outsourced (subcontracting) with a higher cost. Each item is produced through a multi-stage production process in one or several production facilities. Each production stage can include parallel workstations (or machines). Each machine has a limited capacity and a minimum capacity utilisation restriction per period. Due to negligible set-up times, production occurs without set-up costs. Further, the product-machine assignment matrix is given. Finished products are stored in a centralised warehouse, which has a limited storage capacity. Each finished product has a period-specific minimum inventory level over the planning horizon. The finished products are transported from the facilities to this centralised warehouse. Transportation between the warehouse and the facilities has been outsourced to a third party that provides ample transportation capacity for an agreed fixed cost. Consequently, transportation and its cost is not included by the APP. Figure 3.1 illustrates a simple example where six machines are distributed over two facilities, two products are produced at two stage production processes and each

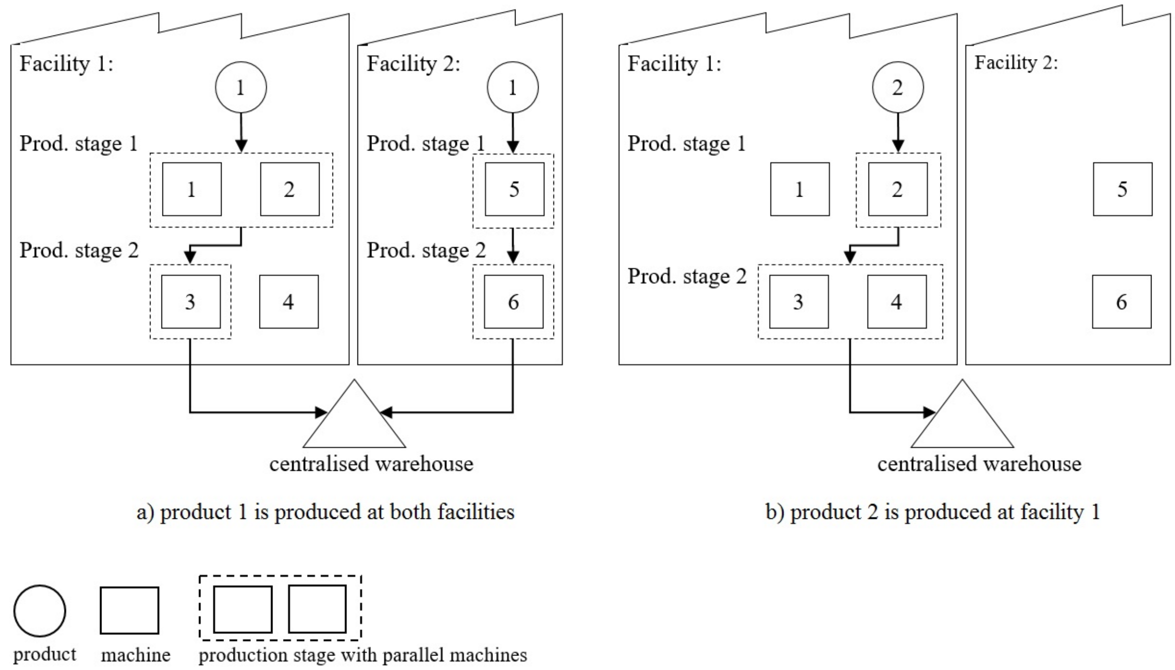


Figure 3.1: Multi-item multi-facility capacity planning problem with parallel machines.

production stage includes a maximum of two parallel machines. The following flexibility instruments are considered.

Shift planning

The production capacity is determined by working time, which is determined by the selection of a shift model. A shift model is defined by the number and length of shifts within a time period and usually limited by external regulations and laws (Sillekens et al., 2011). Each machine runs a certain shift model which can only be changed at the beginning of a period in order to adapt capacity requirements. The cost of a shift model in the APP is usually determined from the cost of workforce, which is generally not affected by the production quantity but rather by the number and the length of a shift model. The cost for increasing or decreasing the workforce level is included in the cost of changing a shift model. For further details about flexible shift planning, we refer to Askar and Zimmermann (2007).

Overtime account

We introduce an overtime account per machine to store overtime hours along the planning horizon. The maximum of overtime hours which can be stored in an overtime account per period depends on the selected shift model. Overtime hours stored in the overtime account are paid according to the cost of workforce in overtime. Furthermore, the overtime account includes an upper and a lower boundary in compliance with regulations and collective agreements. There is a surcharge cost, if overtime hours exceed the upper boundary. The lower boundary specifies the maximum number of hours from the overtime account which can be used to balance the under-times or vacation times during regular working time.

Flexible maintenance

A set of preventive maintenance activities is given for each machine. If a machine is under maintenance, the production on the machine is interrupted, i.e., maintenance requires 100% of the machine's capacity. Further, we assume that the maintenance activity must be repeated after a fixed time interval. The main challenge is to schedule these activities in such a way that they are optimally carried out during periods when machines are underutilised.

3.3 Model Formulation

We introduce the model formulation based on the conventional definition of production and inventory quantities as decision variables. The complete notation is summarised in Table 3.1.

The objective function (3.1) minimises total holding costs, subcontracting costs, fixed costs of the shift model, direct costs of the workforce in overtime and the costs of the workforce from the overtime account.

$$
\begin{aligned}
&\min C = \\
&\underbrace{\sum_{t\in T}\sum_{i\in I} h_i \cdot y_{ti}}_{\text{holding cost}} \quad + \quad \underbrace{\sum_{t\in T}\sum_{i\in I} c_i^{os} \cdot z_{ti}}_{\text{subcontracting}} \quad + \quad \underbrace{\sum_{t\in T}\sum_{s\in S}\sum_{l\in L} c_{tsl}^{fs} \cdot \varphi_{tsl}}_{\text{fixed cost of shift model}} \quad + \\
&\quad \underbrace{\sum_{t\in T}\sum_{l\in L} c_l^{fc} \cdot \delta_{tl}}_{\text{cost of changing shift model}} \quad + \quad \underbrace{\sum_{t\in T}\sum_{l\in L} \zeta_{tl}^{-} \cdot k_l}_{\text{direct cost of workforce in overtime}} \quad + \quad \underbrace{\sum_{t\in T}\sum_{l\in L} \xi_{tl} \cdot k_l \cdot r_l}_{\text{cost of workforce from overtime account}} . \qquad (3.1)
\end{aligned}
$$

Table 3.1: Notation.

Sets:	
I	set of products, $I = \{1, ..., I^{max}\}$
T	set of time periods, $T = \{1, ..., T^{max}\}$
L	set of workstations, $L = \{1, ..., L^{max}\}$
S	set of shift models, $S = \{1, ..., S^{max}\}$
F	set of facilities, $F = \{1, ..., F^{max}\}$
H_l	set of maintenance activities at workstation $l \in L$, $H_l = \{1, ..., H_l^{max}\}$
G_i	set of production stages for product $i \in I$, $G_i = \{1, ..., G_i^{max}\}$
L_{ig}	set of workstations per product $i \in I$ available at production stage $g \in G_i$, $L_{ig} = \{1, ..., L_{ig}^{max}\}$
L_f	set of workstations at facility $f \in F$, $L_f = \{1, ..., L_f^{max}\}$
I_l	set of products per workstation $l \in L$, $I_l = \{1, ..., I_l^{max}\}$
Parameters	
K_{tsl}^N	available working time at workstation l on shift model s in period t
K_{tsl}^O	available overtime at workstation l on shift model s in period t
a_{il}	production capacity consumption per product i at workstation l
d_{ti}	demand of product i in period t
m_{tl}	minimum production capacity utilisation at workstation l in period t (in %)
s_{ti}	safety stock level of product i in period t
W_t	storage capacity of finished goods in period t (in palettes)
h_i	storage cost of product i per unit per period
e_i	quantity of product i per palette
b_i	initial inventory level of product i at the beginning of the planning horizon
c_i^{os}	subcontracting cost per unit of product i
c_{tsl}^{fs}	cost of running shift model s at workstation l in period t
c_l^{fc}	cost of changing a shift model at workstation l
n_{lh}	production capacity consumption by maintenance activity h at workstation l
o_{lh}	interval of maintenance activity h (in time periods) at workstation l
c_l^{fsh}	workforce per-hour cost of regular working time at workstation l, identical over all workstations within each facility
k_l	workforce cost of overtime at workstation l
r_l	coefficient to calculate the supplement time to be paid for surcharges of overtime at workstation l
p_l^{mv}	maximum vacation time from the overtime account at workstation l
p_l^{ml}	maximum level of overtime account at workstation l

Decision variables:

x_{tsil}	quantity of product i produced at workstation l on shift model s in time period t
z_{ti}	quantity of product i subcontracted in period t
y_{ti}	inventory of product i at the end of period t
φ_{tsl}	indicator if workstation l runs on shift model s in period t $(= 1)$, otherwise $(= 0)$
δ_{tl}	indicator if there is a shift change at workstation l at the beginning of period t $(= 1)$, otherwise $(= 0)$
γ_{tsl}	indicator if overtime is used at workstation l on shift model s in period t $(= 1)$, otherwise $(= 0)$
ζ^+_{tl}	unused production capacity of workstation l in period t (in hours)
ζ^-_{tl}	used overtime hours at workstation l in period t
η_{tl}	overtime account balance at workstation l at the end of period t
u_{tslh}	indicator if maintenance activity h is carried out at workstation l with shift model s in period t $(= 1)$, otherwise $(= 0)$
θ_{tl}	overtime reduction level to compensate vacation hours during regular working time at workstation l in period t
ξ_{tl}	exceeded amount from the upper boundary of the overtime account associated to the workforce cost in overtime (in hours)

subject to

- Inventory balance equation.
  Equations (3.2) specify that the inventory level of every product i at the end of every period t is determined by subtracting demand of that period from the sum of the initial inventory and the production in that period. Since we assume zero Work In Process (WIP) inventory, the production quantity of each product in each period is equal to the production quantities at all production stages. The production quantity of an item at a production stage is $\sum_{l\in L_{ig}} \sum_{s\in S} x_{tsil}$. The initial inventory at the beginning of the planning horizon (y_{0i}) is given by b_i.

$$y_{ti} = y_{t-1i} + \sum_{l\in L_{ig}} \sum_{s\in S} x_{tsil} + z_{ti} - d_{ti}. \qquad \forall t \in T, i \in I, g \in G_i \quad (3.2)$$

- Capacity constraints.
  Constraints (3.3) impose production capacity restrictions. The available production capacity of every workstation l in every period t with each shift model s is determined

by subtracting the production capacity consumption of the maintenance activities carried out in that period ($u_{tslh} = 1$) from the sum of the available production capacity of the shift model and overtime.

$$\sum_{i \in I_l} a_{il} \cdot x_{tsil} \leq K^N_{tsl} \cdot \varphi_{tsl} + K^O_{tsl} \cdot \gamma_{tsl} - n_{lh} \cdot u_{tslh}. \qquad \forall t \in T, s \in S, l \in L \quad (3.3)$$

- Flexible maintenance constraints.
  Constraints (3.4) ensure that the maintenance activity h must be carried out at least once during the interval of o_h over the planning horizon. Constraints (3.5) guarantee that maintenance activities are done at workstations under an active shift model.

$$\sum_{t \leq k \leq t+o_{lh}} \sum_{s \in S} u_{kslh} \geq 1, \qquad \forall h \in H_l, t \in \{1, ..., T^{max} - o_{lh}\} \quad (3.4)$$

$$u_{tslh} \leq \varphi_{tsl}. \qquad \forall t \in T, s \in S, l \in L, h \in H_l \quad (3.5)$$

- Minimum production capacity utilisation constraints.
  Constraints (3.6) impose a minimum production capacity utilisation for every workstation l with shift model s in every period t.

$$\sum_{i \in I_l} a_{il} \cdot x_{tsil} \geq K^N_{tsl} \cdot \varphi_{tsl} \cdot m_{tl}. \qquad \forall t \in T, s \in S, l \in L \quad (3.6)$$

- Shift planning constraints.
  Constraints (3.7) ensure that overtime hours can only happen on active shift models. Changes of shift models are determined by constraints (3.8) and (3.9), indicating that only one shift model can run per workstation during one period. Constraints (3.10) and (3.7) together ensure that subcontracting is only available when the highest possible capacity is in use, i.e., a shift system with the highest capacity (shift system S^{max}) plus the corresponding overtime.

$$\gamma_{tsl} \leq \varphi_{tsl}, \qquad \forall t \in T, s \in S, l \in L \quad (3.7)$$

$$\delta_{tl} \geq \varphi_{tsl} - \varphi_{t-1sl} \quad (\varphi_{0sl} = 0), \qquad \forall t \in T, s \in S, l \in L \quad (3.8)$$

$$\sum_{s \in S} \varphi_{tsl} = 1, \qquad \forall t \in T, l \in L \quad (3.9)$$

$$z_{ti} \leq \gamma_{tS^{max}l} \cdot d_{ti}. \qquad \forall t \in T, i \in I, l \in L_i \quad (3.10)$$

- Linking of production stages.

  Due to zero WIP inventory, constraints (3.11) ensure that production quantities of an item at two consecutive production stages must be identical within each facility. The term $l \in L_{ig} \cap L_f$ indicates available machines at production stage g within facility f to work on product i.

$$\sum_{s \in S} \sum_{l \in L_{ig} \cap L_f} x_{tsil} = \sum_{s \in S} \sum_{l \in L_{ig+1} \cap L_f} x_{tsil}. \qquad \forall t \in T, i \in I, g \in G_i \backslash G_i^{max}, f \in F \quad (3.11)$$

- Safety stock and storage constraints of finished goods.

  Safety stock and storage capacity of finished goods are given in constraints (3.12) and (3.13), respectively.

$$y_{ti} \geq s_{ti}, \qquad \forall t \in T, i \in I \quad (3.12)$$

$$\sum_{i \in I} (y_{ti}/e_i) \leq W_t. \qquad \forall t \in T \quad (3.13)$$

- Overtime account constraints.

  Constraints (3.14) determine the unused production capacity under regular working time (ζ_{tl}^{+}) and the overtime hours (ζ_{tl}^{-}) on workstation l in period t. Equations (3.15) determine the overtime account balance at the end of each period by adding overtime hours to the overtime account balance from the previous period and subtracting the reduction time (θ_{tl}) and paid overtime hours from the overtime account (ξ_{tl}). Constraints (3.16) ensure that the reduction of the overtime account (θ_{tl}) in every period t must be smaller than the unused production capacity in the regular working time of that period. Constraints (3.17) impose an upper boundary (p_{tl}^{mv}) for the reduction of the overtime account in every period. Constraints (3.18) enforce an upper boundary for the working time account balance. The zero working time account balance at the end of the planning horizon ($\eta_{T^{max}l} = 0$) ensures that all overtime hours in the overtime account are paid at the end of the planning horizon.

$$\zeta_{tl}^{+} - \zeta_{tl}^{-} =$$
$$\sum_{s \in S} K_{tsl} \cdot \varphi_{tsl} - \sum_{i \in I_l} \sum_{s \in S} a_{til} \cdot x_{tsil} - \sum_{s \in S} \sum_{h \in H_l} n_{lh} \cdot u_{tslh}, \quad \forall t \in T, l \in L \qquad (3.14)$$

$$\eta_{tl} = \eta_{t-1l} + \zeta^-_{tl} - \theta_{tl} - \xi_{tl}, \quad (\eta_{0l} = 0) \qquad \forall t \in T, \forall l \in L \tag{3.15}$$
$$\theta_{tl} \leq \zeta^+_{tl}, \qquad \forall t \in T, l \in L \tag{3.16}$$
$$\theta_{tl} \leq p^{mv}_l, \qquad \forall t \in T, l \in L \tag{3.17}$$
$$\eta_{tl} \leq p^{ml}_l. \qquad (\eta_{T^{max}l} = 0) \qquad \forall t \in T \backslash T^{max}, \forall l \in L \tag{3.18}$$

- Non-negativity and binary decision variables.

$$x_{tsil} \geq 0, \qquad \forall t \in T, s \in S, i \in I, l \in L_i \tag{3.19}$$
$$z_{ti}, y_{ti} \geq 0, \qquad \forall t \in T, i \in I \tag{3.20}$$
$$\theta_{tl}, \eta_{tl}, \xi_{tl}, \zeta^+_{tl}, \zeta^-_{tl} \geq 0, \qquad \forall t \in T, l \in L \tag{3.21}$$
$$u_{tslh} \in \{0, 1\}, \qquad \forall t \in T, s \in S, l \in L, h \in H_l \tag{3.22}$$
$$\delta_{tl} \in \{0, 1\}, \qquad \forall t \in T, l \in L \tag{3.23}$$
$$\varphi_{tsl}, \gamma_{tsl} \in \{0, 1\}. \qquad \forall t \in T, s \in S, l \in L \tag{3.24}$$

3.4 Computational Results

In this section we perform numerical analyses to quantify the cost benefits of the flexibility instruments under different scenarios. Further, we are interested in the tractability of the proposed model in terms of computational effort by commercial and non-commercial MIP solvers. We first discuss results based on a real-world case study, and then present a full factorial design based on randomly generated data to analyse different scenarios. For confidential reasons, we disguised data in the case study.

3.4.1 Case Study

A high volume Electronics Manufacturing Services (EMS) company operates in two production facilities ($F^{max} = 2$). The planning horizon is 12 months ($T^{max} = 12$). At the aggregate level, the company produces six product groups ($I^{max} = 6$) on two-stage production lines ($G^{max} = 2$). Some production stages run with parallel machines. There are ten workstations or machines ($L^{max} = 10$) in total. The minimum capacity utilisation of each machine is set to 30 percent of the available production capacity in every period depending on the selected shift model. Each machine has flexibility to run on two different shift models ($S^{max} = 2$). A shift lasts eight hours and the first shift model consists of three shifts per working day, whereas the second shift model consists of two shifts per working day. The available regular working time of each shift model is determined based on the working calendar. Available overtimes in the first and second shift models are 24 and 16 hours per week, respectively. The corresponding fixed cost of each shift model is determined based on the required workforce level and per-hour cost of regular working time. The per-hour cost of regular working time or overtime in facility two is 20% higher than in facility one. The cost of changing a shift model is specified by the cost of changing the workforce level. There is one type of maintenance activity per machine ($H_l^{max} = 1$), which is carried out every three months ($o_{lh} = 3$). A complete summary on the parameters of the case study is shown in Table 3.5 on page 30.

To evaluate the cost-saving potential of all flexibility instruments and their combinations, we introduce eight different cases as summarised in Table 3.2. Case 8 represents the model presented in Section 3.3 under full flexibility. In all other cases, we assume less flexibility by fixing the variables and conditions of the flexibility instruments not included in the respective case.

Table 3.2: Flexibility instruments in cases.

	Shift planning (3.7)-(3.10)	Overtime account (3.14)-(3.18)	Maintenance (3.4)-(3.5)
Case 1	-	-	-
Case 2	√	-	-
Case 3	-	√	-
Case 4	-	-	√
Case 5	√	√	-
Case 6	√	-	√
Case 7	-	√	√
Case 8	√	√	√

Case 1 represents the inflexible case with a fixed shift model, fixed overtime hours and a predetermined maintenance plan. The constraints kept from Section 3.3 in Case 1 are: Inventory balance constraints (3.2), production capacity constraints (3.3), minimum capacity utilization constraints (3.6), multi-stage production constraints (3.11), safety stock and storage of finished goods constraints (3.12) and (3.13), and non-negativity constraints (3.19)–(3.20). All other constraints are not activated in Case 1. An overview on the constraints activated in the model if a particular flexibility instrument is considered is shown in Table 1.

If flexible shift planning is not included in the model (Cases 1, 3, 4 and 7), we assume that each machine uses the first shift model (three shifts system). If overtime account is not considered in the model (Cases 1, 2, 4 and 6), we assume that available overtime hours are fixed in each period. When overtime hours are used, then the fixed cost of the whole overtime hours will occur in that period, which is determined by multiplying the overtime hours of the first shift model (K^O_{t1l}) with the workforce cost of overtime (k_{tl}) plus the surcharge cost ($r_{tl} \cdot K^O_{t1l} \cdot k_{tl}$) given in (3.25):

$$c^{fo}_{t1l} = K^O_{t1l} \cdot k_{tl} + r_{tl} \cdot K^O_{t1l} \cdot k_{tl}. \qquad \forall t \in T, l \in L \quad (3.25)$$

In cases where flexible maintenance is not included in the model (Cases 1, 2, 3 and 5), we assume that maintenance activities are predetermined and carried out in the first period and repeated every three periods. Therefore, the last term in constraints (3.3), i.e., $n_{lh} \cdot u_{tslh}$, uses predetermined values for these cases. Finally, we assumed that in all eight cases, the minimum capacity utilisation is not affected by the maintenance activities.

Cost analysis of the flexibility instruments

The optimal values of the objective functions for the defined cases are summarised in Table 3.3. Introducing the flexibility instrument shift planning to the inflexible model results in significant cost savings. In this case, workstations can run on an inexpensive shift model whenever they are underutilised. By introducing the overtime account in Case 3, the total production cost only slightly reduces, as the fixed shift model (three shift system) does not require temporary adaptations of working time, because the production capacity levels are sufficient over all periods. Flexible maintenance without shift planning also rarely reduces total production costs. Summarising, the flexibility instruments overtime account and flexible maintenance only lead to significant savings if combined with shift planning.

Computational performance of the proposed models

We test the computational performance of the proposed model using the MIP solver FICO Xpress Optimizer 64-bit v27.01.02 and the open-source solver Cbc 2.7.7 with Google OR-tool[1]. All computations were executed on a 64-bit platform with an Intel Core(TM) i7-4770 3.40 GHz Processor with 32 GB RAM. The run time limit was set to 3600 seconds and the optimality gap to 0.00%. A summary of the computational results is given in

[1]Available at https://code.google.com/p/or-tools/ (accessed May 31, 2018)

Table 3.3: Total production cost improvement (in %) due to flexibility instruments of the case study.

	Optimal value of obj. function (euro)	Cost savings to case 1 (%)
Case 1: Inflexible model	3,661,905	-
Case 2: Shift planning	3,138,980	14.28
Case 3: Overtime	3,661,350	0.02
Case 4: Maintenance	3,657,050	0.13
Case 5: Shift planning & Overtime	3,080,410	15.88
Case 6: Shift planning & Maintenance	3,071,120	16.13
Case 7: Overtime & Maintenance	3,657,050	0.13
Case 8: Shift planning & Overtime & Maintenance	3,036,960	17.07

Table 3.4 reporting the runtime, number of branch and bound nodes, and number of simplex iterations. The run time to solve the MILP varies significantly between the cases. In more detail, the introduction of flexible shift planning to the inflexible model (Case 2) or a combination of flexible shift planning with other flexibility instruments increases the number of simplex iterations executed in the Linear Program (LP) relaxation, the number of branch and bound nodes searched to reach the optimal solution and consequently the computation times due to the introduction of binary variables of flexible shift planning. Case 6, where flexible shift planning and maintenance are combined, results in the highest runtime, number of branch and bound nodes and simplex iterations. Moreover, adding the overtime account (Case 8) reduces the run time, number of branch and bound nodes and simplex iterations. Comparing the performance of Xpress with Cbc, we observe, as expected, that Xpress outperforms Cbc. However, the computational performance is still satisfactory with Cbc for this real problem instance.

3.4.2 Results for Randomly Generated Instances

In order to quantify the cost-savings of the flexibility instruments and the computational performance of the proposed mathematical formulation in different scenarios, we conduct a numerical study based on randomly generated input parameters.

Random instance generator

Taking the input parameters from the case study discussed in the previous section into account, the random instances are generated and drawn mainly by varying the values

Table 3.4: Computational performance of the proposed models of case study.

	Xpress				Cbc			
Case	Time (sec)	Gap (%)	Nodes	Iterations	Time (sec)	Gap (%)	Nodes	Iterations
1	0.00	0.00	1	208	0.00	0.00	1	1
2	5.76	0.00	16635	46020	171	0.00	24201	819996
3	0.00	0.00	1	227	0.00	0.00	1	1
4	0.02	0.00	1	387	0.00	0.00	1	1
5	1.17	0.00	479	12601	17	0.00	979	93820
6	18.89	0.00	57907	152429	440	0.00	58901	2824227
7	0.00	0.00	1	251	0.00	0.00	1	1
8	4.90	0.00	2599	45457	139	0.00	5569	742716

of the following four input parameters: Available regular working times, demand, labour cost and shift change cost. Based on the data set of the case study, the input parameters are either fixed to or varied or drawn from a discrete uniform distribution $UD(a, b)$ on the interval $[a, b]$ and from a normal distribution $N(\mu, \sigma)$ with mean μ and standard deviation σ, respectively. An overview of the main sets and input parameters of the case study and how and in which interval the instances are generated is given in Table 3.5.

For all generated instances we fix the sizes of the numbers of products, time periods, workstations, shift models, maintenance activities, production stages and the distribution of the workstations over the facilities according to the case study values.

We introduce three capacity scenarios. In the first scenario, we assume a high production capacity level by increasing the original working times of the case study by 25%. In the second scenario, we assume a medium production capacity level and keep production capacity unchanged and in the third capacity scenario, we assume a low production capacity level by decreasing the original working times by 25%. The corresponding available overtimes are changed in the same way.

For each category of the generated instances with high, medium and low production capacity levels, we define two demand variation scenarios with low and high demand variation levels. Demand of a product is generated based on the normal distribution with a mean value of the mean demand of the product over the planning horizon from the data-set of the case study and coefficient of variations (CV) of 0.1 for the low demand variation level and 0.4 for the high demand variation level.

For each production capacity and demand level, we define the three scenarios of low, medium and high shift cost levels. As the per-hour cost of regular working time differs between facilities, we assume values of 25, 50 and 100 for the per-hour cost of working time for the low, medium and high shift cost levels, respectively, in facility one. Consequently, we assume values of 30, 60 and 120 for the per-hour cost of working time for the low, medium, and high shift cost levels, respectively, in facility two.

For every combination of production capacity, demand and shift cost level, we define the three scenarios of low, medium and high cost levels of changing a shift model. In the low shift changing cost level, the cost of changing the shift model is assumed to be zero. In the case of the medium shift changing cost level, in accordance with the original data-set, we assume that the shift changing cost is equal to half of the average cost of the cheapest shift model over the planning horizon (indicated by $c_l^{fs,ave}$). Finally, in the high shift changing cost setting, we assume that the shift changing cost is equal to the average cost of the cheapest shift model over the planning horizon. Since the cost of the workforce for overtime is dependent on the per-hour cost of regular working time, it is also updated accordingly.

Initial inventory depends on the average demand per product i over T (d_i^{ave}) and is generated randomly from $UD(d_i^{ave}, 2 \cdot d_i^{ave})$. The safety stock level is calculated as 25% of the sum of the demand in the next two periods. For the final two periods, we consider the demand of the final period T^{max}. All other parameter values remain unchanged.

Cost analysis and computational performance

We randomly generate seven instances for each combination of the four varying parameters in a total of 378 ($= 3*2*3*3$) test instances. Every generated instance is used for testing the defined cases by solving them with Xpress and Cbc. In accordance with Section 3.4.1, the run time limit was set to 3600 seconds and the optimality gap to 0.00%.

Table 3.6 shows the average cost reductions, the corresponding standard deviations, and the minimum and maximum values of the cost reductions (in %) for Cases 2 to 8 in comparison to Case 1 over all instances within each defined scenario. We observe significant cost reductions under flexible shift planning (Cases 2, 5, 6 and 8), especially if production capacity levels are high, shift cost levels are high and shift change costs are low. Under flexible shift planning, expensive and inexpensive shift models can be adapted based on required capacities, which decreases the total cost substantially. However, the cost reduction due to the overtime account and the flexible maintenance (Cases 3, 4 and 7) is very low in the mentioned scenarios.

For the scenario with low production capacity levels, however, we see that a combina-

Table 3.5: Overview of sets and parameters of case study and randomly generated instances.

Set /Parameter	Case study Ranges $[min, max]$	Generated instances Ranges $[min, max]$
Sets		
I	$\{1, ..., 6\}$	$\{1, ..., 6\}$
T	$\{1, ..., 12\}$	$\{1, ..., 12\}$
L	$\{1, ..., 10\}$	$\{1, ..., 10\}$
S	$\{1, 2\}$	$\{1, 2\}$
H	$\{1\}$	$\{1\}$
G	$\{1, 2\}$	$\{1, 2\}$
Shift system		
K^N_{tsl}	$l \in L_1$, $s = 1 : [402, 514]$, $s = 2 : [250, 338]$ $l \in L_2$, $s = 1 : [330, 516]$, $s = 2 : [202, 340]$	low: 80%, medium: 100%, high :120% of original K^N_{tsl}
a_{il}	$[0.207, 0.545]$	$[0.207, 0.545]$
c^{fsh}_l,k_l	$l \in L_1 : [50]$ $l \in L_2 : [60]$	$l \in L_1$: low: 25, medium: 50, high: 100 $l \in L_2$: low: 30, medium: 60, high: 120
c^{fs}_{tsl}	$s = 1 : [24765, 120805]$ $s = 2 : [16510, 80537]$	$c^{fsh}_{tl} \cdot K^N_{tsl}$
c^{fc}_l	$[\frac{c^{fs,ave}_{1l}}{2}, \frac{c^{fs,ave}_{1l}}{2}]$	high: $c^{fs,ave}_{1l}$, medium: $\frac{1}{2}c^{fs,ave}_{1l}$, low: 0
Overtime		
K^o_{tsl}	$l \in L_1$, $s = 1 : [72, 120]$, $s = 2 : [48, 80]$	low:80%, medium:100%, high: 120% of original K^o_{tsl}
K^o_{tsl}	$l \in L_2$, $s = 1 : [72, 120]$, $s = 2 : [48, 80]$	
p^{mv}_l	$[120, 120]$	$[120, 120]$
p^{ml}_l	$[120, 120]$	$[120, 120]$
r_l	$[1.5, 1.5]$	$[1.5, 1.5]$
c^{fo}_{tsl}	$(1 + r_{tl}) \cdot K^o_{tsl} \cdot k_{tl}$	$(1 + r_{tl}) \cdot K^o_{tsl} \cdot k_{tl}$
Inventory		
b_i	$[5000, 133205]$	$UD(d^{ave}_i, 2 \cdot d^{ave}_i)$
s_{ti}	$[15000, 80000]$	$\frac{d_{t+1i}+d_{t+2i}}{4}$ $(t + 1 \geq T^{max} \rightarrow d_{T^{max}i})$
e_i	$[480, 672]$	$[480, 672]$
W_t	$[1800, 1800]$	$[1800, 1800]$
h_i	$[0.12, 0.28]$	$[0.12, 0.28]$
Maintenance		
n_{lh}	$[48, 48]$	$[48, 48]$
o_{lh}	$[3, 3]$	$[3, 3]$
Demand and others		
d_{ti}	$[6000, 180920]$	$N(d^{ave}_i, CV \cdot d^{ave}_i)$: low $CV = 0.1$, high $CV = 0.4$
m_{tl}	$[0.3, 0.3]$	$[0.3, 0.3]$
c^{os}_i	$[T^{max} \cdot \max\{h_j \| j \in I\}, T^{max} \cdot \max\{h_j \| j \in I\}]$	$[T^{max} \cdot \max\{h_j \| j \in I\}, T^{max} \cdot \max\{h_j \| j \in I\}]$

tion of all three flexibility instruments (Case 8) becomes more important. On average, the cost saving increases by nearly 4% in Case 8 compared to Case 2 in the low production capacity scenario. In contrast to this, the average cost saving potential from Case 2 to Case 8 is 0.76% in the scenario with high production capacity levels. The cost reductions due to overtime account and flexible maintenance increase as production capacity levels decrease. This highlights the importance of integrating all three flexibility instruments into the production planning process, especially when production capacity is limited. If the variation in demand increases, the cost reduction under flexible shift planning mostly decreases. A high variability in demand generally implies frequent shift changes which increasing cost, hence, decreases the cost reduction compared to the situation with low demand variation. However, overtime account and flexible maintenance lead to higher cost reductions in case of higher demand variations by capturing demand fluctuations to avoid frequent shift changes.

Table 3.7 presents a summary of the computational efficiency when solving the proposed mathematical formulation with Xpress and with Cbc. Taking the solver Xpress into account, we observe that all instances in Cases 1, 3, and 7 are solved to optimality within the given run time limit. The average runtime increases by using the solver Cbc. However, Cbc is also able to solve most of the instances to optimality. For the sake of completeness, Table 3.8 presents the detailed runtime performance of Cbc relative to Xpress (runtime Cbc)/(runtime Xpress). The runtime of Cbc relative to Xpress increases, especially when capacity levels are low.

3.5 Conclusions

In this chapter, we introduced an MILP for a multi-item multi-facility multi-stage capacity planning problem with parallel machines by taking the flexibility instruments shift planning, overtime account and flexible maintenance into consideration. To analyse the impact of such factors as available production capacity, demand variation, shift cost and shift change cost on the cost-saving potential of the flexibility instruments, computational results both for a real-world case study and randomly generated test instances were provided. Our computational experiments gave insights into the value of integrating the flexibility instruments into the APP individually and in combination with each other, which helps us with understanding the interaction effects of the flexibility instruments and their respective effectiveness. We observed that the integration of flexible shift planning into production capacity planning has the most substantial effect on the total cost

Table 3.6: Total production cost improvement.

	Case 2		Case 3		Case 4		Case 5		Case 6		Case 7		Case 8	
	Mean (Stdv) (%)	Min Max (%)	Mean (Stdv) (%)	Min Max (%)	Mean (Stdv) (%)	Min Max (%)	Mean (Stdv) (%)	Min Max (%)	Mean (Stdv) (%)	Min Max (%)	Mean (Stdv) (%)	Min Max (%)	Mean (Stdv) (%)	Min Max (%)
Capacity														
low	2.22	0.00	0.81	0.04	2.71	0.89	3.39	0.70	5.32	1.98	5.52	2.32	6.17	2.39
	(1.34)	6.18	(0.64)	2.93	(1.76)	8.77	(1.40)	8.12	(2.11)	13.94	(2.27)	12.48	(2.32)	15.12
medium	16.13	7.84	0.12	0.00	0.20	0.00	16.93	10.08	17.30	10.19	0.27	0.00	17.82	11.87
	(3.15)	24.33	(0.29)	1.90	(0.55)	5.50	(2.87)	24.67	(2.96)	25.27	(0.67)	6.29	(2.74)	25.34
high	25.68	17.83	0.01	0.00	0.02	0.00	26.03	19.29	26.22	19.28	0.02	0.00	26.44	19.91
	(2.85)	31.34	(0.08)	0.63	(0.11)	1.00	(2.68)	31.36	(2.73)	31.55	(0.16)	1.63	(2.63)	31.55
Demand														
low	14.96	0.00	0.21	0.00	0.85	0.00	15.60	0.88	16.40	2.23	1.78	0.00	16.82	2.87
	(10.52)	30.70	(0.39)	2.06	(1.47)	7.66	(10.16)	30.71	(9.53)	31.02	(2.76)	11.89	(9.27)	31.02
high	14.11	0.00	0.44	0.00	1.15	0.00	15.03	0.70	15.91	1.98	2.21	0.00	16.56	2.39
	(9.75)	31.34	(0.66)	2.93	(1.80)	8.77	(9.37)	31.36	(8.66)	31.55	(3.12)	12.48	(8.39)	31.55
Shift cost level														
low	12.65	0.00	0.40	0.00	1.29	0.00	13.63	1.42	14.57	2.35	2.25	0.00	15.24	3.04
	(8.87)	26.04	(0.60)	2.93	(2.13)	8.77	(8.51)	26.50	(7.67)	26.42	(3.35)	12.48	(7.42)	26.73
medium	14.66	0.13	0.34	0.00	1.03	0.00	15.45	0.70	16.28	1.98	2.04	0.00	16.84	2.39
	(10.15)	28.97	(0.59)	2.76	(1.59)	5.93	(9.78)	29.01	(9.05)	29.17	(3.00)	10.70	(8.77)	29.21
high	15.90	0.00	0.26	0.00	0.74	0.00	16.55	0.79	17.30	2.17	1.76	0.00	17.74	2.52
	(11.00)	31.34	(0.44)	2.19	(0.96)	3.79	(10.65)	31.36	(10.18)	31.55	(2.41)	7.82	(9.94)	31.55
Shift change cost														
low	15.71	0.87	0.34	0.00	1.11	0.00	16.46	2.37	17.48	4.07	2.18	0.00	17.97	4.78
	(10.22)	31.34	(0.52)	2.93	(1.84)	8.77	(9.80)	30.71	(8.99)	31.55	(3.17)	12.48	(8.69)	31.55
medium	14.27	0.11	0.31	0.00	0.91	0.00	15.10	0.88	15.83	2.23	1.88	0.00	16.40	2.87
	(10.15)	31.34	(0.53)	2.70	(1.50)	7.61	(9.80)	31.36	(9.18)	31.55	(2.81)	11.89	(8.92)	31.55
high	13.85	0.00	0.27	0.00	0.86	0.00	14.65	0.70	15.37	1.98	1.79	0.00	15.91	2.39
	(9.97)	28.99	(0.60)	2.93	(1.58)	8.27	(9.62)	28.99	(9.00)	29.27	(2.85)	12.48	(8.76)	29.27
All Instances	14.53	0.00	0.33	0.00	1.00	0.00	15.32	0.70	16.15	1.98	2.00	0.00	16.69	2.39
	(10.17)	31.34	(0.55)	2.39	(1.65)	8.77	(9.79)	31.36	(9.12)	31.55	(2.96)	12.48	(8.85)	31.55

Table 3.7: Summary of computational performance: Instances solved to optimality, avg. runtime of instances solved to optimality and avg. optimality gap of instances not solved to optimality within 1 hour.

	Xpress			Cbc		
Case	Optimality (%)	AvgTime (sec)	AvgGap (%)	Optimality (%)	AvgTime (sec)	AvgGap (%)
1	100	0.38	0.00	99.61	17.45	0.01
2	91.29	111.98	0.23	60.95	431.44	0.14
3	100	0.02	0.00	99.74	0.05	0.00
4	99.74	11.77	0.12	62.27	233.49	0.01
5	97.63	28.60	0.11	79.42	248.18	0.06
6	75.46	190.20	0.18	32.45	1,153.19	0.23
7	100	0.03	0.00	86.28	0.10	0.00
8	92.88	58.69	0.07	66.23	338.30	0.07

reduction. However, if the production capacity is very limited, the implementations of the overtime account and flexible maintenance also result in considerable cost reductions. We found that a combination of all flexibility instruments are crucial for manufacturers that face high demand variabilities, especially when the available production capacity levels are low.

The numerical experiments illustrated the computational efficiency of the proposed model and its applicability for similar real-world capacity planning problems. In terms of the computational performance, the solver Cbc was generally able to solve the problem instances in an acceptable time and with an acceptable optimality gap although outperformed by the solver Xpress. Hence, the proposed production planning model can serve as an effective decision support for real problem instances in order to balance available capacities over a mid-term planning horizon.

Acknowledgements

The presented research work in this chapter was accomplished under project KoKa | Development of a decision support system for capacity adjustments at optimal costs within the scope of the Intelligent Production program from the Austrian Research Promotion Agency (FFG).

Table 3.8: Runtime of Cbc relative to Xpress.

		Shift	shift	Cases							
Capa.	Demand	chan. cost	cost	1	2	3	4	5	6	7	8
high	low	low	low	2.60	18.16	2.99	2.12	46.59	41.12	4.00	28.96
			med.	2.95	3.59	2.94	2.83	6.32	6.46	3.53	16.21
			high	2.37	4.34	2.83	2.29	6.12	6.43	4.56	8.22
		med.	low	3.34	23.25	6.08	3.75	34.68	13.08	4.00	10.49
			med.	3.10	4.68	2.71	2.48	7.30	8.71	4.23	9.55
			high	2.47	4.45	2.58	2.64	6.19	6.38	3.45	6.59
		high	low	2.41	14.47	3.13	2.62	18.93	18.00	3.73	20.53
			med.	2.47	4.59	3.17	2.46	5.98	8.04	3.26	7.34
			high	4.06	7.32	5.03	3.25	7.69	9.15	4.84	8.23
	med.	low	low	3.10	28.47	3.75	3.20	75.75	30.27	4.62	32.53
			med.	2.93	13.95	3.84	3.40	7.31	15.23	4.54	11.78
			high	2.94	6.11	3.52	3.44	8.49	8.89	4.02	7.41
		med.	low	3.84	26.13	3.52	4.04	32.61	47.54	5.70	43.39
			med.	4.99	19.35	5.81	5.52	16.58	15.98	6.80	17.53
			high	2.78	8.81	3.17	2.80	6.80	11.84	4.27	10.93
		high	low	5.81	33.74	3.37	5.48	27.49	25.54	3.97	21.66
			med.	3.52	6.48	3.29	4.80	12.03	10.48	3.94	12.62
			high	5.32	6.83	4.71	4.08	9.51	10.18	5.51	8.31
med.	low	low	low	4.93	2.41	2.48	3.59	21.73	1.00	3.29	5.77
			med.	5.01	47.50	3.74	3.79	12.78	24.75	5.30	19.74
			high	3.65	17.84	4.11	4.62	13.88	14.20	4.09	20.63
		med.	low	3.49	1.27	1.04	0.94	20.82	1.00	2.95	1.16
			med.	7.37	39.48	6.64	6.17	22.96	37.90	7.99	32.94
			high	7.35	18.01	5.44	5.25	12.15	13.69	6.22	30.30
		high	low	4.09	0.93	0.74	0.48	3.76	1.00	4.88	0.99
			med.	5.48	21.00	5.01	4.17	16.80	22.76	5.08	17.96
			high	5.96	24.27	6.03	5.45	13.98	11.99	5.70	26.32
	med.	low	low	18.85	16.04	5.10	5.71	93.28	1.29	4.51	26.47
			med.	12.77	41.01	5.19	5.64	18.07	43.39	5.65	26.98
			high	2.53	19.85	3.30	2.10	11.87	21.83	3.26	11.00
		med.	low	4.39	8.33	3.24	2.95	60.79	2.09	3.04	17.00
			med.	4.10	31.54	3.71	2.39	17.14	26.78	3.64	21.27
			high	5.05	11.83	3.61	3.66	6.86	18.72	3.20	12.69

Table 3.8: Runtime of Cbc relative to Xpress.

		Shift	shift	Cases							
Capa.	Demand	chan. cost	cost	1	2	3	4	5	6	7	8
med.	med.	high	low	3.57	18.92	3.75	4.46	40.36	7.67	4.90	18.80
			med.	4.37	16.62	3.08	3.00	14.62	18.77	3.66	15.85
			high	3.69	9.45	2.95	3.78	12.05	17.24	4.13	15.69
low	low	low	low	38.90	15.91	3.37	68.02	91.42	1.61	2.29	143.65
			med.	17.94	8.04	4.59	65.50	13.93	3.71	3.55	11.47
			high	14.85	3.29	4.33	60.44	8.90	10.21	4.24	14.18
		med.	low	26.83	5.54	4.78	19.65	19.47	1.08	2.98	37.30
			med.	20.51	9.14	4.22	79.05	13.68	5.01	3.18	22.49
			high	68.78	8.03	5.50	81.48	11.80	7.99	4.32	9.43
		high	low	38.99	9.27	4.63	44.49	36.02	1.00	4.71	46.42
			med.	46.24	8.74	6.54	146.86	13.78	3.97	4.76	28.15
			high	48.83	12.55	6.47	124.10	11.37	9.22	5.69	13.31
	med.	low	low	11.28	9.80	5.88	166.46	22.91	23.05	9.39	27.64
			med.	5.92	75.13	3.59	20.30	8.06	23.05	3.21	8.86
			high	23.93	20.38	3.81	31.00	9.38	18.29	5.37	5.70
		med.	low	18.03	3.63	3.63	37.76	47.60	2.85	6.99	18.01
			med.	12.94	11.57	3.98	105.46	12.46	12.45	6.13	15.42
			high	64.52	9.60	5.09	48.33	10.83	11.45	7.66	8.88
		high	low	10.16	17.57	4.91	44.04	20.75	2.82	8.38	45.76
			med.	12.52	22.31	4.90	52.50	12.21	16.04	6.61	18.50
			high	20.33	11.02	5.78	29.11	8.49	17.07	7.20	8.92

Chapter 4

Planning Approaches for Stochastic Capacitated Lot-Sizing with Service Level Constraints

In this chapter, we investigate a stochastic capacitated lot-sizing problem. An optimal solution of this problem requires the integration of dynamic safety stock planning into lot-sizing. We first present a stochastic dynamic program. Then, we introduce a new integrated mixed-integer linear program with service-level constraints. The integrated model endogenously sets dynamic safety stocks over non-equidistant lengths of replenishment cycles. Since there is limited capacity, soft service-level constraints are introduced to guarantee a feasible solution. The integrated model is extended to also account for re-planning opportunities under rolling horizon planning. In the experimental study, we compare the performance of the integrated model to the stochastic dynamic program and the widely-used sequential approaches. We observe that, if available capacity increases, the integrated model closes the gap to the theoretical lower bound that has been obtained from the stochastic dynamic program. We find that, if capacity is limited, the integrated model outperforms the sequential approaches because it yields identical service levels with lower inventories. However, we show that, if there is sufficient flexibility (capacity) under rolling horizon planning, the integrated model must also take re-planning opportunities into account in order to avoid the build-up of excess safety stock.

4.1 Introduction

Stochastic capacitated lot-sizing finds cost-optimal decisions regarding when and how much of a product must be produced in order to satisfy uncertain demand in the presence of limited capacity. This problem can be found in a wide variety of disciplines and industrial sectors and plays a key role in Manufacturing Resource Planning (MRP II) and Advanced Planning Systems (APS).

One of the main characteristics of stochastic capacitated lot-sizing is demand uncertainty. A common way of dealing with demand uncertainty is to hold safety stocks by solving a safety stock planning problem. Lot-sizing and safety stock planning problems are usually solved sequentially (see, e.g., de Kok and Fransoo, 2003). When using planning software, such as MRP II and APS, it is common to first solve a safety stock planning problem in order to capture the uncertainty by using *exogenous safety stocks* and then solve a deterministic lot-sizing problem that takes exogenous safety stocks into account as a constraint.

The main pitfall of such a sequential approach is that it does not take the connection between the decisions on safety stocks and lot-sizing into account. To be more precise, safety stocks are placed to buffer against the demand uncertainty over *replenishment cycles*. The replenishment cycle depends on the production plan, which is the timespan between two consecutive production periods. The production plan clearly impacts the safety stock requirement. On the other hand, the safety stock requirement also affects the production planning problem, because the safety stock quantities need to be produced at some point. Therefore, the two problems are closely interrelated.

Despite this major shortcoming, the main advantage of the sequential approaches is that they reduce the complexity caused by the use of an integrated model that addresses both problems simultaneously, known as the *stochastic lot-sizing problems*. Schneeweiss (2003) refers to this type of decomposition as a *constructional distributed decision making* system where a complex operations research problem (here, stochastic lot-sizing problem) is divided into more manageable sub-problems. The use of separate approaches to determine safety stocks (a higher level) and to produce plans (a lower level) reduces the complexity of the original stochastic lot-sizing problem.

The stochastic lot-sizing problem has recently attracted increasing attention. The existing literature makes considerable contributions by proposing the use of integrated models to address this problem (see Sections 2.2 and 2.3 for related literature). As an integrated model, a chance-constrained program (mainly followed by a deterministic equivalent formulation with service-level constraints) is usually proposed in order to place *dynamic safety stocks* over replenishment cycles. In chance-constrained programs, constraints can be violated with a prescribed probability (Charnes and Cooper, 1959). More specifically, target inventory levels at the beginning of replenishment cycles are determined based on a certain probability that stochastic demand is satisfied.

So far, however, the main aspects of a stochastic lot-sizing problem have rarely been addressed in an integrated model concurrently: The existing literature either neglects capacitated cases, assumes exogenous replenishment cycles, or focuses on production times and disregards production quantities. From a modelling perspective, an endogenous determination of replenishment cycles has not yet been fully solved. The existing literature mainly uses a complete enumeration of every possible length of replenishment cycles. Moreover, in the context of limited production capacity, the common hard service-level constraints are not realistic assumptions, as they can lead to infeasibility if the inventory level is insufficient (Albey et al., 2015).

In practice, dynamic lot-sizing is predominantly implemented under a rolling horizon approach. This means that, based on a forecasted demand, a finite planning horizon problem is initially solved. The production plan is implemented during a number of initial periods. The planning horizon is rolled forward and, at the same time, forecasted demand is updated as more information is gained over time and the system status is updated as the actual demand is realised. The main advantage of a rolling horizon approach is that it provides *re-planning opportunities* for the solution of periods that have not yet been implemented. In other words, based on the realised demand and updated forecasted demand, there are opportunities to adjust the original plan (re-plan) over the course of the following (rolled) planning. Another shortcoming of the integrated models we found in the literature is that they ignore the re-planning opportunities that exist under rolling horizon planning.

Both the sequential and the integrated approaches are decision generators for prescribing lot-sizes and safety stocks (Schneeweiss, 2003). Thus, the problem of lot-sizing with uncertain demand under rolling horizon planning can be more appropriately addressed by using a *stochastic dynamic program* (see, e.g., Mula et al. (2006)). A Stochastic Dynamic Program (SDP) enables us to anticipate the value of the re-planning opportunities that exist during the next stages (periods) as early as during the current stage (period). However, the SDPs usually suffer from the curse of dimensionality. In particular, industrial lot-sizing models often have large problem sizes and various production constraints with real-world settings (Jans and Degraeve, 2008), all of which makes it very difficult to solve them optimally by using a stochastic dynamic program.

Compared to an SDP, using a chance-constrained program can substantially reduce the size and complexity of the integrated model, but this is at the expense of neglecting the future re-planning opportunities. Thus, integrating re-planning options into the chance-constrained programs can transform them into more realistic ways of addressing the underlying stochastic optimisation problem.

To this end, we introduce the following three Mixed-Integer Linear Programs (MILPs), each of which addresses the underlying stochastic optimisation problem with its own degree of comprehensiveness and complexity.

(i) Sequential approach.

We develop an MILP to address the deterministic counterpart of the original problem where random demand is replaced by forecasted demand and assumed to be deterministic. Demand uncertainty is captured by using a simple method for calculating exogenous safety stocks.

(ii) Integrated model.

The integrated model enables the endogenous placement of dynamic safety stocks during replenishment cycles. This model is presented as a new MILP based on a chance-constrained program.

(iii) Integrated model with re-planning opportunities.

The aforementioned integrated model does not take the possibility of re-planning as it is found under a rolling horizon approach into account. In order to address this issue within an MILP, we introduce a simple heuristic and include it in the integrated model.

We summarise the contributions of our work as follows:

- We propose a new MILP that endogenously determines the length of the replenishment cycles observed in different types of stochastic capacitated lot-sizing problems. Our general approach enables the use of different methods for endogenously placing dynamic safety stocks, e.g., the use of service-level constraints. We further introduce soft service-level constraints to prevent infeasibility brought on by the capacity restrictions in case of an insufficient inventory.

- In order to interpolate the non-linear order-up-to-levels (target inventory levels) in the underlying MILP, we propose a bivariate linearisation technique with an improved triangulation method.

- We develop a new procedure to evaluate the total costs of the MILPs with respect to an identical realised service-level under rolling horizon planning. Our experimental investigation shows that, in the presence of limited capacity, re-planning opportunities under rolling horizon planning are restricted, and hence placement of dynamic safety stocks is beneficial. However, if capacity is not binding, demand uncertainty beyond the re-planning periods (the time that elapses between the starting points of the two consecutive rolling schedules) can be accommodated through additional production, which reduces the need for dynamic safety stocks. This is how a sequential approach with exogenous safety stock calculations can outperform the integrated model with excess dynamic safety stocks. We propose a simple approach that anticipates the upcoming re-planning opportunities under rolling horizon planning. The heuristic reduces the order-up-to-levels with respect to the available safety capacity. The new version of the integrated model outperforms the sequential approach.

- Based on small test instances, we evaluate the absolute performance of the proposed integrated model by means of a stochastic dynamic program. We observe that the integrated model can get nearer to the theoretical lower bound if the available capacity increases.

This chapter is structured as follows: Section 4.2 describes the Stochastic Capacitated Lot-Sizing Problem (S-CLSP). Section 4.3 presents an SDP. Section 4.4 introduces three MILPs, i.e., a sequential approach, an integrated model with service level-constraints, and a re-planning-opportunity adjusted integrated model. Section 4.5 presents the experimental design and the numerical results. Conclusions and further research are presented in the final section.

4.2 Problem Description

We assume a lot-sizing problem where final products with random demand are produced directly from raw materials by a single machine with limited capacity. We find cost-optimal decisions regarding when and how much of a product is produced in order to satisfy random demand under a prescribed target cycle service-level. We use the following general notation style. A set is indicated by a capital letter in a calligraphic font, e.g., $\mathcal{T} = \{1, 2, .., |\mathcal{T}|\}$. Random variables are denoted with the use of a hat notation, e.g., $\hat{d}_{tp}$. Decision variables are indicated by letters in bold font. The vector of the elements of a decision variable or a parameter is created by dropping one or more indices, e.g., $Y_t = (Y_{tp})_{\forall p \in \mathcal{P}}$ indicates the vector Y_t over all products in every period $t \in \mathcal{T}$.

Consider a finite planning horizon with $|\mathcal{T}|$ equidistant and discrete time periods. We have $|\mathcal{P}|$ products and every product $p \in \mathcal{P} = \{1, 2, ..., |\mathcal{P}|\}$ has a demand $\hat{d}_{tp}$ in every period t, which is an exogenous, non-stationary and mutually independent random variable that follows a continuous demand distribution with mean d_{tp} and standard deviation σ_{tp}. A production order can be issued for product p at time t and become immediately available within that period after a fixed setup time (z_p). The production capacity needed for producing a unit of product p is k_p and the total available production capacity at time t is denoted by K_t. There can be a minimum production quantity ($q_p^{\min}$) whenever we produce product p. We allow production orders with lower quantities than $q_p^{\min}$, but the per unit shortfall is penalised with a cost coefficient c_p^{minq}. Every product p can have an initial inventory at the beginning of the planning horizon (y_{0p}). The left-over inventory is carried to the following periods and a linear inventory holding cost (c_p^{h}) is incurred per period and per unit of product p. Table 4.1 presents the notation.

We set the setup costs to zero and assume that setup times represent setup costs. As discussed by Kang et al. (2014), while the formulation of lot sizing problems with setup costs is traditional, it gets complicated when it comes to estimating the setup costs in a logical manner. Setup costs will depend on the opportunity cost of the capacity reserved for setups, which is not known until an optimal solution to the problem has been found. Since we do not take setup costs into account, we use minimum production quantities to obtain lot-sizes. In practice, these constraints are fairly common (Jans and Degraeve, 2008).

The unsatisfied demand is backordered. We limit the amount of the backorder quantities through the use of a target cycle fill-rate (β^c-service level). The cycle fill-rate is defined as the fraction of demand during the replenishment cycle that is met immediately from stock and production (see, e.g., Tempelmeier, 2011). The replenishment cycle is the time that elapses between two consecutive production periods. Without loss of generality, we assume that t is a production period and let τ_t denote the first production period after period t (with $\tau_{|\mathcal{T}|} = |\mathcal{T}|{+}1$). The timespan between the start of period t and the end of period $\tau_t - 1$ describes a replenishment cycle. With the β^c-service level, we ensure that

Table 4.1: Notation.

Sets	
$\mathcal{P} = \{1, ..., \|\mathcal{P}\|\}$	set of products
$\mathcal{T} = \{1, ..., \|\mathcal{T}\|\}$	set of periods
Parameters	
$\hat{d}_{tp}$	random demand of product p in period t
d_{tp}	mean (forecasted) demand of product p in period t
σ_{tp}	standard deviation of demand of product p in period t
τ_t	production period following period t
c_p^{h}	inventory holding cost of product p per unit and per period
c_p^{bl}	backorder penalty cost for product p per unit and per period
c_p^{minq}	penalty cost for per unit shortfall of the min. production level of product p
z_p	setup time for product p
K_t	capacity(time) available in period t
k_p	capacity consumption(time) to produce one unit of product p
y_{0p}	initial inventory of product p at the beginning of a planning horizon
q_p^{min}	minimum production quantity for product p
β^c	target cycle fill-rate

the ratio of the cumulative demand during a replenishment cycle that cannot be satisfied immediately is less than $1 - \beta^c$.

4.3 Stochastic Dynamic Program

Let $S_t = (Y_{tp}, B_{tp}, D_{tp})_{\forall p \in \mathcal{P}}$ denote the state of the system at time t, where Y_{tp} indicates the inventory level for product p, and B_{tp} and D_{tp} denote the cumulative backordered quantity and the cumulative demand of product p during the current replenishment cycle up to time t. At every state S_t, we decide whether or not to place an order with a magnitude of $\mathbf{q}_t \in \mathcal{Q}_t^\star$, where $\mathcal{Q}_t^\star$ indicates the feasible action space. The decision $\mathbf{q}_t \in \mathcal{Q}_t^\star$ at state S_t is determined by using a policy function $Q_t^\pi(S_t)$ $\forall \pi \in \Pi$, where Π indicates the set of all possible policies. The function $Q_t^\pi(S_t)$ produces a feasible decision $\mathbf{q}_t$ by using policy π at state S_t.

The model of the system dynamics is represented by $S_{t+1} = S^{\mathrm{M}}(S_t, \mathbf{q}_t, \hat{d}_{t+1})$, where $S^{\mathrm{M}}(.)$ refers to the transition function. $S^{\mathrm{M}}(.)$ implies that, if we are at state S_t and then make decision $\mathbf{q}_t$, the random demand $\hat{d}_{t+1}$, which is observed during time $t+1$, will transfer the state of the system to S_{t+1}. More specifically, the transition function of each state variable of the system is presented as follows. The inventory level is given by $Y_{t+1} = Y_t + \mathbf{q}_t - \hat{d}_{t+1}$. The transition function of the cumulative demand is written as $D_{t+1} = 1_{\{\mathbf{q}_t=0\}} \cdot D_t + \hat{d}_{t+1}$, where $1_{\{true\}} = 1$ and otherwise $= 0$. The transition function of the cumulative backordered quantities can be mathematically introduced as follows.

$$B_{t+1} = 1_{\{\mathbf{q}_t>0\}} \cdot \left[\hat{d}_{t+1} - [Y_t + \mathbf{q}_t]^+\right]^+ + 1_{\{\mathbf{q}_t=0\}} \cdot \left(B_t + \left[\hat{d}_{t+1} - [Y_t]^+\right]^+\right), \tag{4.1}$$

where $[x]^+ = max(x, 0)$. The first term determines the new backorder quantities that occur in the first period of the replenishment cycle after production. The second term sums up the backorder quantities that newly occur in all periods after the first period within the replenishment cycle.

The feasible action space for the vector $\mathbf{q}_t$ is tightened according to the limited capacity and the target cycle fill-rate as follows.

- Limited capacity.

  The production quantity of product p at time t ($\mathbf{q}_{tp}$) is limited by the available production capacity. Mathematically,

$$k_p \cdot \mathbf{q}_{tp} \leq K_t - \sum_{p' \in \mathcal{P} | p' \neq p} \left(k_{p'} \cdot \mathbf{q}_{tp'} + 1_{\mathbf{q}_{tp'}>0} \cdot z_{p'}\right). \tag{4.2}$$

  This implies that the feasible region for $\mathbf{q}_{tp}$ will depend on K_t, k_p, z_p and the production quantities of other products $\mathbf{q}_{tp'}$ with $p' \neq p$.

- Target cycle fill-rate.

  Following the definition of the β^c-service level, we ensure that the ratio between the cumulative backorder quantities and the cumulative demand at time t must be less than $1 - \beta^c$. Mathematically, we can write

$$\frac{B_t}{D_t} \leq 1 - \beta^c. \tag{4.3}$$

  Note that this definition is equivalent to the definition of the cycle fill-rate under stationary cases, where the expected backorder quantities in every cycle relates to the average replenishment quantity (Silver and Bischak, 2011, Tempelmeier, 2011). The feasible space action will further depend on B_t, D_t and the β^c-service level.

The following cost function quantifies the inventory holding cost and the cost for any shortfall of the minimum production quantity at state S_t.

$$C_t(S_t, \mathbf{q}_t) = \sum_{p \in \mathcal{P}} \left(c_p^{\mathrm{h}} \cdot [Y]_{tp}^{+} + c_p^{\mathrm{minq}} \cdot [q_p^{\mathrm{min}} - \mathbf{q}_{tp}]^{+} \right). \tag{4.4}$$

Note that constraints (4.2) and (4.3) may contradict each other. In other words, we may not have sufficient capacity to obtain the prescribed cycle fill-rate. Thus, constraints in (4.2) and (4.3) may lead to an infeasible action space.

Similar to what Powell (2014) proposed, the optimal policy for the optimisation problem can be characterised by solving the following Bellmann equation.

$$Q_t^{\star}(S_t) = \operatorname*{argmin}_{\mathbf{q}_t \in \mathcal{Q}_t^{\star}} \left(C(S_t, \mathbf{q}_t) + \mathbb{E} \left\{ \sum_{t'=t+1}^{|\mathcal{T}|} C(S_{t'}, Q_{t'}^{\pi}(S_{t'})) \middle| S_t \right\} \right), \tag{4.5}$$

where $S_{t+1} = S^{\mathrm{M}}(S_t, \mathbf{q}_t, \hat{d}_{t+1})$ and $\mathbf{q}_t = Q_t^{\pi}(S_t) \in \mathcal{Q}_t^{\star}$ $\forall \pi \in \Pi$.

Equation (4.5) implies that an optimal policy must be designed in such a way that it generates feasible decisions of $\mathbf{q}_t$ that minimise the total cost of the current state (the first term) and the expected total cost of all the following stages (the second term). However, in a stochastic dynamic program, the service level constraints (4.3) can complicate the solution of equation (4.5). An alternative way of solving the problem is to use a backorder-penalty cost minimisation approach instead of a service level approach. In this case, the cumulative backorder quantities in every state S_t are penalised in the cost function with a per unit backorder penalty cost (c_p^{bl}). The state of the system is further reduced to $S_t = (Y_{tp}, B_{tp})_{\forall p \in \mathcal{P}}$. Moreover, in the cost function (4.4), we further add a backorder penalty cost term as follows.

$$C_t(S_t, \mathbf{q}_t) = \sum_{p \in \mathcal{P}} \left(c_p^{\mathrm{h}} \cdot Y_{tp}^{+} + c_p^{\mathrm{bl}} \cdot B_{tp} + c_p^{\mathrm{minq}} \cdot [q_p^{\mathrm{min}} - \mathbf{q}_{tp}]^{+} \right). \tag{4.6}$$

4.4 Mixed-Integer Linear Programs

4.4.1 Sequential Approach

The sequential approach divides the S-CLSP into a safety stock planning problem and a deterministic Capacitated Lot-Sizing Problem (CLSP). First, a safety stock planning problem is solved by applying a simple rule-of-thumb method in order to compute exogenous safety stocks. For example, certain days of supply are usually guaranteed by the use of a specific amount of the forecasted demand as safety stock. Then, a deterministic CLSP, where the forecasted demand is assumed to be deterministic, is solved by taking exogenous safety stocks as constraints.

The backlogged demand of product p at the end of each period is penalised with the per unit penalty cost coefficient c_p^{bl} in the objective function. Soft constraints are introduced for the product-specific minimum production quantity level ($q_p^{\min}$), as well as for a final inventory level at the end of the planning horizon (y_p^{fi}) and for the safety stock (y_{tp}^{sst}) at the end of every period. These soft constraints ensure the feasibility of the production plans by allowing shortfalls of $q_p^{\min}$, y_p^{fi} and y_{tp}^{sst} if the capacity is not sufficient. The corresponding shortfalls are penalised in the objective function with different cost coefficients. Note that the final inventory level is used for preventing the truncated horizon effect. Table 4.2 presents the new notation for the sequential approach.

Table 4.2: Notation for the sequential approach.

Parameters	
c_p^{sst}	penalty cost for a unit of shortfall of the safety stock level of product p per period
$c_p^{y^{\mathrm{fi}}}$	penalty cost for a unit of shortfall of the final inventory level of product p
y_{tp}^{sst}	exogenous safety stock level for product p at the end of period t
y_p^{fi}	final inventory level for product p at the end of a planning horizon
Decision variables	
$\mathbf{q}_{tp}$	production quantity of product p in period t
$\boldsymbol{\delta}_{tp}$	setup indicator for product p in period t
$\mathbf{y}_{tp}^{+}$	inventory on-hand for product p at the end of period t
$\mathbf{y}_{tp}^{-}$	backlogged quantity for product p at the end of period t
$\boldsymbol{\zeta}_{tp}$	shortfall of minimum production quantity of product p in period t
$\boldsymbol{\psi}_{tp}$	shortfall of the exogenous safety stock level of product p in period t
$\boldsymbol{\rho}_p$	shortfall of the final inventory level of product p

The MILP is given below.

$$\min \; TC = \sum_{p \in \mathcal{P}} \sum_{t \in \mathcal{T}} \left(c_p^{\mathrm{h}} \cdot \mathbf{y}_{tp}^{+} + c_p^{\mathrm{bl}} \cdot \mathbf{y}_{tp}^{-} + c_p^{\mathrm{minq}} \cdot \boldsymbol{\zeta}_{tp} + c_p^{\mathrm{sst}} \cdot \boldsymbol{\psi}_{tp} \right) + \sum_{p \in \mathcal{P}} c_p^{y^{\mathrm{fi}}} \cdot \boldsymbol{\rho}_p, \tag{4.7}$$

subject to

$$\mathbf{y}_{tp}^{+} - \mathbf{y}_{tp}^{-} = y_{0p} + \sum_{i \in \mathcal{T} | i \leq t} (\mathbf{q}_{ip} - d_{ip}), \qquad \forall t \in \mathcal{T}, p \in \mathcal{P} \tag{4.8}$$

$$\sum_{p \in \mathcal{P}} (k_p \cdot \mathbf{q}_{tp} + z_p \cdot \boldsymbol{\delta}_{tp}) \leq K_t, \qquad \forall t \in \mathcal{T} \tag{4.9}$$

$$k_p \cdot \mathbf{q}_{tp} \leq K_t \cdot \boldsymbol{\delta}_{tp}, \qquad \forall t \in \mathcal{T}, p \in \mathcal{P} \tag{4.10}$$

$$\mathbf{q}_{tp} \geq \boldsymbol{\delta}_{tp} \cdot q_p^{\min} - \boldsymbol{\zeta}_{tp}, \qquad \forall t \in \mathcal{T}, p \in \mathcal{P} \tag{4.11}$$

$$y_{0p} + \sum_{i \in \mathcal{T} | i \leq t} (\mathbf{q}_{ip} - d_{ip}) \geq y_{tp}^{\mathrm{sst}} - \boldsymbol{\psi}_{tp}, \qquad \forall t \in \mathcal{T} | t < |\mathcal{T}|, p \in \mathcal{P} \tag{4.12}$$

$$\mathbf{y}_{|\mathcal{T}|,p}^{+} - \mathbf{y}_{|\mathcal{T}|,p}^{-} \geq y_p^{\mathrm{fi}} - \boldsymbol{\rho}_p, \qquad \forall p \in \mathcal{P} \tag{4.13}$$

$$\mathbf{y}_{tp}^{+} \geq 0, \mathbf{y}_{tp}^{-} \geq 0, \qquad \forall t \in \mathcal{T}, p \in \mathcal{P} \tag{4.14}$$

$$\mathbf{q}_{tp} \geq 0, \boldsymbol{\delta}_{tp} \in \{0, 1\}, \boldsymbol{\zeta}_{tp} \geq 0, \boldsymbol{\psi}_{tp} \geq 0, \qquad \forall t \in \mathcal{T}, p \in \mathcal{P} \tag{4.15}$$

$$\boldsymbol{\rho}_p \geq 0. \qquad \forall p \in \mathcal{P} \tag{4.16}$$

The objective function (4.7) minimises the inventory holding costs, backlog penalty costs and shortfall costs of the minimum production quantities and safety stocks, as well as the final inventory level. The inventory balance equations are given in (4.8). Constraints (4.9) limit the capacity. The setup indicator logic is presented in (4.10). The soft constraints for the minimum production quantity, the safety stocks and the final inventory levels are given in (4.11), (4.12) and (4.13), respectively. Note that, in the final period $|\mathcal{T}|$, we drop the soft safety stock constraints (4.12), because we use the constraints on the final inventory (4.13). If the safety stock in the last period is higher than the final inventory level, we set the final inventory level to the safety stock. The binary and non-negative decision variables are given in (4.14), (4.15) and (4.16).

4.4.2 Integrated Model

We first introduce the chance-constrained program for the integrated model on which the MILP is built.

Chance-constrained program

Let $\mathcal{T}^\star$ denote a set of all production periods over the planning horizon, i.e., $\mathcal{T}^\star = \{\forall i \in \mathcal{T} | \sum_{p\in\mathcal{P}} \mathbf{q}_{ip} > 0\}$. Moreover, let $\hat{D}_{tp}$ $\forall t \in \mathcal{T}^\star$ denote the random replenishment cycle demand, which is determined as the cumulative random demand between t and $\tau_t - 1$, i.e., $\hat{D}_{tp} = \hat{d}_{tp} + ... + \hat{d}_{\tau_t-1,p} = \sum_{i\in\mathcal{R}_t} \hat{d}_{ip}$, where $\mathcal{R}_t = \{i \in \mathcal{T} | t \leq i < \tau_t\}$. Note that the time representation of index t is slightly different from the one defined in the stochastic dynamic program in Section 4.3, specifically, demand in period t is $\hat{d}_{tp}$ instead of $\hat{d}_{t+1,p}$.

We use $\hat{\mathbf{y}}^-_{tp} \geq 0$ to indicate the random backlogged quantity of product p at the end of period $t \in \{0, 1, ..., |\mathcal{T}|\}$, where $\hat{\mathbf{y}}^-_{0p} \geq 0$ shows the initial backlogged quantity. In order to limit the amount of cumulative backorder quantities between t and $\tau_t - 1$, we introduce the following chance-constraint with respect to a target cycle fill-rate.

$$\frac{\mathbb{E}\left\{[\hat{\mathbf{y}}^-_{tp} - [\hat{\mathbf{y}}^-_{t-1p} - \mathbf{q}_{tp}]^+]^+ + \sum_{i=t+1}^{\tau_t-1}[\hat{\mathbf{y}}^-_{ip} - \hat{\mathbf{y}}^-_{i-1p}]^+\right\}}{\mathbb{E}\left\{\hat{D}_{tp}\right\}} \leq 1 - \beta^c, \qquad \forall t \in \mathcal{T}^\star, p \in \mathcal{P} \quad (4.17)$$

where $\mathbb{E}$ denotes the expectation operator and $[x]^+ = max(x, 0)$. The first term in the numerator determines backorders that happened in period t directly after the production as the positive difference between the backlogged quantities at the end of period t and at the beginning of period t directly after production. The second term in the numerator sums up the new backorders that happened later than t within the current replenishment cycle. The constraints (4.17) ensure that the ratio of the replenishment cycle demand that is backordered must be limited to $1 - \beta^c$. The complete chance-constrained program can be written as follows.

$$\min\ \mathbb{E}\{\hat{TC}\} = \mathbb{E}\left\{\sum_{p\in\mathcal{P}}\sum_{t\in\mathcal{T}}\left(c^{\mathrm{h}}_p \cdot \hat{\mathbf{y}}^+_{tp} + c^{\mathrm{minq}}_p \cdot \boldsymbol{\zeta}_{tp}\right) + \sum_{p\in\mathcal{P}} c^{y^{\mathrm{fi}}}_p \cdot \boldsymbol{\rho}_p\right\}, \quad (4.18)$$

subject to

$$\hat{\mathbf{y}}^+_{tp} - \hat{\mathbf{y}}^-_{tp} = y_{0p} + \sum_{i\in\mathcal{T}|i\leq t}(\mathbf{q}_{ip} - \hat{d}_{ip}), \qquad \forall t \in \mathcal{T}, p \in \mathcal{P} \quad (4.19)$$

$$\frac{\mathbb{E}\left\{[\hat{\mathbf{y}}^-_{tp} - (\hat{\mathbf{y}}^-_{t-1p} - \mathbf{q}_{tp})^+]^+ + \sum_{i=t+1}^{\tau_t-1}(\hat{\mathbf{y}}^-_{ip} - \hat{\mathbf{y}}^-_{i-1p})^+\right\}}{\mathbb{E}\left\{\hat{D}_{tp}\right\}} \leq 1 - \beta^c, \qquad \forall t \in \mathcal{T}^\star, p \in \mathcal{P} \quad (4.20)$$

$$\hat{\mathbf{y}}^+_{|\mathcal{T}|,p} - \hat{\mathbf{y}}^-_{|\mathcal{T}|,p} \geq y^{\mathrm{fi}}_p - \boldsymbol{\rho}_p, \qquad \forall p \in \mathcal{P} \quad (4.21)$$

$$\mathbf{q}_{tp} \geq 0, \boldsymbol{\delta}_{tp} \in \{0,1\}, \boldsymbol{\zeta}_{tp} \geq 0, \qquad \forall t \in \mathcal{T}, p \in \mathcal{P} \quad (4.22)$$
$$(4.9), (4.10), (4.11), (4.14), (4.16).$$

The objective function (4.18) minimises the expected total cost, including the random inventory holding cost and costs of not meeting the minimum production quantity and the final inventory level. Equations (4.19) indicate the random inventory level at the end of every period t for p. The chance-constraints are given in (4.20). The soft constraints for the random final inventory level are presented in (4.21). The other constraints, which are identical to the MILP for the sequential approach, are: The upper bound on capacity (4.9), the setup indicator logic (4.10), soft constraints on minimum production quantities (4.11), and the binary and non-negative decision variable constraints (4.14), (4.16) and (4.22).

Mixed-integer linear program

Let $\mathbf{b}_{tp}$ and $\mathbf{m}_{tp}$ denote the mean and variance of the cumulative demand from t to $\tau_t - 1$, i.e., $\hat{D}_{tp}$, respectively. Mathematically, $\mathbf{b}_{tp} = \sum_{i=t}^{\tau_t - 1} d_{ip}$ and $\mathbf{m}_{tp} = \sum_{i=t}^{\tau_t - 1} \sigma_{ip}^2$. As has been shown by Charnes and Cooper (1959), we can write the chance-constraint in (4.17) for every $t \in \mathcal{T}^\star$ as follows where $\mathbf{b}_{tp} = \mathbb{E}\{\hat{D}_{tp}\}$.

$$\mathbb{E}\left\{[\hat{\mathbf{y}}_{tp}^- - (\hat{\mathbf{y}}_{t-1p}^- - \mathbf{q}_{tp})^+]^+ + \sum_{i=t+1}^{\tau_t - 1} (\hat{\mathbf{y}}_{ip}^- - \hat{\mathbf{y}}_{i-1p}^-)^+\right\} \leq (1-\beta^c) \cdot \mathbb{E}\left\{\hat{D}_{tp}\right\} \Rightarrow$$
$$\int_{\mathbf{y}_{tp}^{\mathrm{TL}}}^{\infty} (x - \mathbf{y}_{tp}^{\mathrm{TL}}) f_p(x) dx \leq (1-\beta^c) \cdot \mathbf{b}_{tp}. \quad (4.23)$$

In (4.23), $\mathbf{y}_{tp}^{\mathrm{TL}}$ indicates the target inventory level (order-up-to level) immediately after production in period t and $f_p(.)$ denotes the continuous density function of $\hat{D}_{tp}$. In other words, the order-up-to-level at the beginning of the replenishment cycle must be specified in such a way that the amount of the cumulative backorder quantities until the end of period $\tau_t - 1$, which is determined by the first-order loss function $\int_{\mathbf{y}_{tp}^{\mathrm{TL}}}^{\infty} (x - \mathbf{y}_{tp}^{\mathrm{TL}}) f_p(x) dx$, is limited by the β^c-service level, i.e., it is smaller than or equal to $(1-\beta^c) \cdot \mathbf{b}_{tp}$.

By means of $\mathbf{y}_{tp}^{\mathrm{TL}}$, we can transform the chance-constrained program into an MILP with service-level constraints. By using service-level constraints, we ensure that the inventory level at the beginning of a replenishment cycle is equal to or higher than the target inventory level. Note that the service level constraints are only binding at the beginning of a replenishment cycle. This is due to the fact that $\mathbf{b}_{tp}$ and $\mathbf{m}_{tp}$, and consequently $\mathbf{y}_{tp}^{\mathrm{TL}}$, have the highest values at the beginning of a replenishment cycle. Thus, we only apply service-level constraints in production periods or at the beginning of the planning horizon since those are the only times when a replenishment cycle can begin.

If the capacity is limited and the inventory at the beginning of replenishment cycles is not sufficient, service-level constraints may prevent us from finding feasible solutions for the underlying problem. To avoid infeasibility, we need to reduce β^c and consequently $\mathbf{y}_p^{\mathrm{TL}}$. Following the example given by Albey et al. (2015), with soft service-level constraints we endogenously reduce $\mathbf{y}_p^{\mathrm{TL}}$ to avoid infeasibility if the capacity becomes binding. Let $\mathbf{y}_{tp}^{\mathrm{ds}}$ denote the dynamic safety stock level during the replenishment cycle that starts from period t. We know that $\mathbf{y}_{tp}^{\mathrm{TL}} = \mathbf{b}_{tp} + \mathbf{y}_{tp}^{\mathrm{ds}}$ (Silver et al., 2017). The introduction of soft service-level constraints means that we can allow for a shortfall of either $\mathbf{y}_{tp}^{\mathrm{ds}}$, or $\mathbf{b}_{tp}$, or both. Thus, we introduce two continuous decision variables $\boldsymbol{\psi}_{tp}^{\mathrm{ds}} \geq 0$ and $\boldsymbol{\psi}_{tp}^{\mathrm{bl}} \geq 0$ that correspond to the shortfall of the dynamic safety stock part and to the shortfall of the mean of the replenishment cycle demand part of the order-up-to-level, respectively.

A different interpretation of these shortfalls has been given by (Albey et al., 2015). $\boldsymbol{\psi}_{tp}^{\mathrm{bl}}$ represents planned backlog; based on the expected demand ($\mathbf{b}_{tp}$), this is what we plan to backlog in order to accommodate for limited capacity. $\boldsymbol{\psi}_{tp}^{\mathrm{ds}}$, on the other hand, represents the planned deviation from the required dynamic safety stock. The per unit shortfall of the dynamic safety stock is penalised by c_p^{ds}. Moreover, the per unit shortfall of the mean of the replenishment cycle demand is penalised by c_p^{bl}, which, in the deterministic model, is used as the backlog penalty cost coefficient. We assume that the per unit shortfall of the dynamic safety stock is less costly than the per unit shortfall of the mean of the replenishment cycle demand ($c_p^{\mathrm{ds}} < c_p^{\mathrm{bl}}$). This is because a violation of dynamic safety stocks is usually more acceptable than a violation of demand satisfaction. The MILP for the integrated model is introduced as follows. Table 4.3 presents the new decision variables.

$$\min\ TC = \sum_{p\in\mathcal{P}}\sum_{t\in\mathcal{T}}\left(c_p^{\mathrm{h}}\cdot\mathbf{y}_{tp}^{+} + c_p^{\mathrm{minq}}\cdot\boldsymbol{\zeta}_{tp} + c_p^{\mathrm{bl}}\cdot\boldsymbol{\psi}_{tp}^{\mathrm{bl}} + c_p^{\mathrm{ds}}\cdot\boldsymbol{\psi}_{tp}^{\mathrm{ds}}\right) + \sum_{p\in\mathcal{P}} c_p^{y^{\mathrm{fi}}}\cdot\boldsymbol{\rho}_p, \tag{4.24}$$

subject to

$$y_{0p} + \mathbf{q}_{tp} \geq \mathbf{y}_{tp}^{\mathrm{TL}} - \boldsymbol{\psi}_{tp}^{\mathrm{bl}} - \boldsymbol{\psi}_{tp}^{\mathrm{ds}}, \qquad \forall p \in \mathcal{P}, t = 1 \quad (4.25)$$

$$y_{0p} + \sum_{i \in \mathcal{T} | i \leq t} \mathbf{q}_{ip} - \sum_{i \in \mathcal{T} | i < t} d_{ip} \geq \mathbf{y}_{tp}^{\mathrm{TL}} - \boldsymbol{\psi}_{tp}^{\mathrm{bl}} - \boldsymbol{\psi}_{tp}^{\mathrm{ds}} - (1 - \boldsymbol{\delta}_{tp}) \cdot M, \quad \forall t \in \mathcal{T}, p \in \mathcal{P} | t > 1 \quad (4.26)$$

$$\boldsymbol{\psi}_{tp}^{\mathrm{ds}} \leq \mathbf{y}_{tp}^{\mathrm{ds}}, \qquad \forall t \in \mathcal{T},\ p \in \mathcal{P} \quad (4.27)$$

$$\mathbf{y}_{tp}^{\mathrm{TL}} \geq 0, \boldsymbol{\psi}_{tp}^{\mathrm{ds}} \geq 0, \boldsymbol{\psi}_{tp}^{\mathrm{bl}} \geq 0, \qquad \forall t \in \mathcal{T}, p \in \mathcal{P} \quad (4.28)$$

$$(4.8), (4.9), (4.10), (4.11), (4.13), (4.14), (4.15) \text{ and } (4.16).$$

The objective function in (4.24) minimises the total inventory cost, the shortfall costs of the minimum production quantity and the final inventory level, as well as the associated shortfall costs of the order-up-to-level in the soft service-level constraints. The soft service-level constraints in (4.25) and (4.26) are used instead of the chance-constraints in (4.20). (4.25) and (4.26) ensure that the inventory level at the beginning of the first period and at the beginning of the production periods are at least equal to the order-up-to-level. In constraints (4.27), we introduce the upper bound for $\boldsymbol{\psi}_{tp}^{\mathrm{ds}}$, which is the corresponding dynamic safety stock level. Since $\boldsymbol{\psi}_{tp}^{\mathrm{ds}}$ is cheaper than $\boldsymbol{\psi}_{tp}^{\mathrm{bl}}$, $\boldsymbol{\psi}_{tp}^{\mathrm{bl}}$ will only be used if $\boldsymbol{\psi}_{tp}^{\mathrm{ds}}$ reaches the upper bound. Finally, we add the following constraints from the MILP of the sequential approach, which also apply to the integrated model: The inventory balance equations (4.8), the upper bound on capacity (4.9), the setup indicator logic (4.10), soft constraints on minimum production quantity (4.11) and final inventory levels (4.13), as well as the binary and non-negative decision variables constraints (4.14), (4.15) and (4.16). Note that we dropped the soft safety stock constraints (4.12) and their associated costs in the objective function.

So far, we introduced the MILP for the integrated model without discussing how to determine $\mathbf{b}_{tp}$, $\mathbf{m}_{tp}$ and $\mathbf{y}_{tp}^{\mathrm{TL}}$ within an MILP. Since $f_p(.)$ is characterised by $\mathbf{b}_{tp}$ and $\mathbf{m}_{tp}$, $\mathbf{y}_{tp}^{\mathrm{TL}}$ depends on $\mathbf{b}_{tp}$ and $\mathbf{m}_{tp}$, i.e., $\mathbf{y}_{tp}^{\mathrm{TL}}(\mathbf{b}_{tp}, \mathbf{m}_{tp})$. Given β^c, the demand distribution of $\hat{D}_{tp}$, $\mathbf{b}_{tp}$ and $\mathbf{m}_{tp}$, we can easily specify $\mathbf{y}_{tp}^{\mathrm{TL}}$ in (4.23) by using a numerical method. However, $\mathbf{y}_{tp}^{\mathrm{TL}}$ is a bivariate non-linear function. Later in this section, we introduce a linearisation technique that uses a triangulation method as an MILP to approximate $\mathbf{y}_{tp}^{\mathrm{TL}}$.

The main issue that still remains to be dealt with is how to endogenously specify $\mathbf{b}_{tp}$ and $\mathbf{m}_{tp}$, both of which depend on the production periods (the solution of the underlying

Table 4.3: New decision variables for the integrated model.

$\mathbf{y}_{tp}^{TL}$	order-up-to-level of product p in period t
$\mathbf{y}_{tp}^{\mathrm{ds}}$	dynamic safety stock level of product p part of the order-up-to-level in period t
$\boldsymbol{\psi}_{tp}^{\mathrm{ds}}$	shortfall of the dynamic safety stock part of the order-up-to-level of product p in period t
$\boldsymbol{\psi}_{tp}^{\mathrm{bl}}$	shortfall of the mean of the replenishment cycle demand ($\mathbf{b}_{tp}$) part of the order-up-to-level of product p in period t

lot-sizing problem). We address this issue in the following section.

Demand uncertainty parameters $\mathbf{b}_{tp}$ and $\mathbf{m}_{tp}$ over endogenous replenishment cycles

Table 4.4 summarises the new parameters and decision variables introduced in this section.

Table 4.4: New notation for demand uncertainty parameters.

Parameters	
b_{tp}^{max}	upper bound on $\mathbf{b}_{tp}$ for product p in period t
m_{tp}^{max}	upper bound on $\mathbf{m}_{tp}$ for product p in period t
Decision variables	
$\mathbf{b}_{tp}$	mean of the replenishment cycle demand of product p in period t
$\mathbf{m}_{tp}$	variance of the replenishment cycle demand of product p in period t
$\bar{\mathbf{b}}_{tp}$	fraction of $\mathbf{b}_{tp}$ from b_{tp}^{max} for product p in period t
$\bar{\mathbf{m}}_{tp}$	fractional of $\mathbf{m}_{tp}$ from m_{tp}^{max} for product p in period t
$\mathbf{u}_{tp}$	auxiliary decision variable corresponding to $\bar{\mathbf{b}}_{tp}$ for product p in period t
$\mathbf{c}_{tp}$	auxiliary decision variable corresponding to $\bar{\mathbf{m}}_{tp}$ for product p in period t

- Mean of replenishment cycle demand ($\mathbf{b}_{tp}$).

  Let $b_{tp}^{\max}$ denote the upper bound of the mean of the replenishment cycle demand that corresponds to period t, which is obtained by summing up the mean demands from the beginning of period t to the end of the planning horizon.

$$b_{tp}^{\max} = \sum_{i \in \mathcal{T} | i \geq t} d_{ip}. \qquad \forall t \in \mathcal{T}, p \in \mathcal{P} \qquad (4.29)$$

Let the continuous decision variable $\bar{\mathbf{b}}_{tp} = \mathbf{b}_{tp}/b_{tp}^{\max}$ denote the fraction of the mean of the replenishment cycle demand of product p in period t from $b_{tp}^{\max}$. We use the fractional values because they improve the computational performance by avoiding the implementation of bigM formulations. Furthermore, we define the auxiliary continuous decision variable $\mathbf{u}_{tp}$ to track the values of $\bar{\mathbf{b}}_{tp}$ over the planning horizon.

$$\mathbf{u}_{tp} \leq \boldsymbol{\delta}_{tp}, \qquad \forall t \in \mathcal{T}, p \in \mathcal{P} \tag{4.30}$$

$$\mathbf{u}_{tp} \geq \bar{\mathbf{b}}_{tp} - (1 - \boldsymbol{\delta}_{tp}), \qquad \forall t \in \mathcal{T}, p \in \mathcal{P} \tag{4.31}$$

$$\mathbf{u}_{tp} \leq \bar{\mathbf{b}}_{tp} + (1 - \boldsymbol{\delta}_{tp}), \qquad \forall t \in \mathcal{T}, p \in \mathcal{P} \tag{4.32}$$

$$\bar{\mathbf{b}}_{tp} \cdot b_{tp}^{\max} = \sum_{i \in \mathcal{T} | i \geq t} d_{ip} - \sum_{i \in \mathcal{T} | i > t} \mathbf{u}_{ip} \cdot b_{ip}^{\max}, \qquad \forall t \in \mathcal{T}, p \in \mathcal{P} \tag{4.33}$$

$$\mathbf{u}_{tp}, \bar{\mathbf{b}}_{tp} \geq 0. \qquad \forall t \in \mathcal{T}, p \in \mathcal{P} \tag{4.34}$$

Constraints (4.30) force the auxiliary decision variable $\mathbf{u}_{tp}$ to take on the value of zero if t is not a production period ($\boldsymbol{\delta}_{tp} = 0$) otherwise ($\boldsymbol{\delta}_{tp} = 1$), constraints (4.31) and (4.32) force the auxiliary variable $\mathbf{u}_{tp}$ to take on the same value as the fractional mean of the replenishment cycle demand at the beginning of period t ($\bar{\mathbf{b}}_{tp}$). Equations (4.33) calculate the mean of the replenishment cycle demand at the beginning of period t ($\bar{\mathbf{b}}_{tp} \cdot b_{tp}^{\max}$) by summing up the mean demand of the periods, including and following the current micro period, less the mean of the replenishment cycle demand that corresponds to the following replenishment cycles stored by the auxiliary decision variables.

- Variance of replenishment cycle demand ($\mathbf{m}_{tp}$).

  Let $m_{tp}^{\max}$ denote the upper bound of the variance of the replenishment cycle demand that corresponds to period t.

$$m_{tp}^{\max} = \sum_{i \in \mathcal{T} | i \geq t} \sigma_{ip}^2. \qquad \forall t \in \mathcal{T}, p \in \mathcal{P} \tag{4.35}$$

Let the continuous decision variable $\bar{\mathbf{m}}_{tp} = \mathbf{m}_{tp}/m_{tp}^{\max}$ denote the fraction of the variance of the replenishment cycle demand of product p in period t from $m_{tp}^{\max}$. Furthermore, we define the auxiliary decision variable $\mathbf{c}_{tp}$ to track the values of $\bar{\mathbf{m}}_{tp}$ over the planning horizon. As with the determination of the mean of the replenishment cycle demand, the following formulation presents the calculation of the variance of the replenishment cycle demands.

$$\mathbf{c}_{tp} \leq \boldsymbol{\delta}_{tp}, \qquad \forall t \in \mathcal{T}, p \in \mathcal{P} \qquad (4.36)$$

$$\mathbf{c}_{tp} \geq \bar{\mathbf{m}}_{tp} - (1 - \boldsymbol{\delta}_{tp}), \qquad \forall t \in \mathcal{T}, p \in \mathcal{P} \qquad (4.37)$$

$$\mathbf{c}_{tp} \leq \bar{\mathbf{m}}_{tp} + (1 - \boldsymbol{\delta}_{tp}), \qquad \forall t \in \mathcal{T}, p \in \mathcal{P} \qquad (4.38)$$

$$\bar{\mathbf{m}}_{tp} \cdot m_{tp}^{\max} = \sum_{i \in \mathcal{T} | i \geq t} \sigma_{ip}^2 - \sum_{i \in \mathcal{T} | i > t} \mathbf{c}_{ip} \cdot m_{ip}^{\max}, \qquad \forall t \in \mathcal{T}, p \in \mathcal{P} \qquad (4.39)$$

$$\mathbf{c}_{tp}, \bar{\mathbf{m}}_{tp} \geq 0. \qquad \forall t \in \mathcal{T}, p \in \mathcal{P} \qquad (4.40)$$

Table 4.5 shows a numerical example of how the mean and variance of the replenishment cycle demand are determined for product p over a planning horizon with six periods. The production periods are specified by $\boldsymbol{\delta}_{tp} = 1$. For example, the mean of the replenishment cycle demand in period one is $\mathbf{b}_{1p} = 400$, which is the sum of the mean demand from period one to the end of period six less the mean value of the following replenishment cycle, i.e., $\mathbf{u}_{5p} \cdot b_{5p}^{\max} = 200$.

Table 4.5: Mean and variance of replenishment cycle demand.

t	1	2	3	4	5	6
$\boldsymbol{\delta}_{tp}$	**1**	0	0	0	**1**	0
d_{tp}	100	100	100	100	100	100
$\mathbf{b}_{tp}$	400	300	200	100	200	100
$\mathbf{u}_{tp} \cdot b_{tp}^{\max}$	400	0	0	0	200	0
σ_{tp}^2	400	400	400	400	400	400
$\mathbf{m}_{tp}$	1600	1200	800	400	800	400
$\mathbf{c}_{tp} \cdot m_{tp}^{\max}$	1600	0	0	0	800	0

Bivariate linearisation method

We approximate the bivariate non-linear order-up-to-level function ($\mathbf{y}_{tp}^{\mathrm{TL}}$) by a triangulation method. (See, e.g., Vielma et al. (2010) and Rebennack and Kallrath (2015) for more details on triangulation methods). Think of a rectangular region the horizontal and vertical sides of which represent the ranges of the mean and variance of the replenishment cycle demand. We construct an approximation grid by breaking down the region into small rectangles based on predetermined approximation points on the horizontal and vertical sides. Every small rectangle is then further divided into two triangles, i.e., the upper and the lower triangles. The vertices of the triangle, in which the value of $\mathbf{y}_{tp}^{\mathrm{TL}}$ is located, are then used for the linear interpolation.

More specifically, on the x-axis, which represents the values of the mean of the replenishment cycle demand ($\mathbf{b}_{tp}$), we define $|\mathcal{N}|$ predetermined approximation points b_{tpi}, where $b_{tpi} \in [0, b_{tp}^{\max}]$ for every $i = \{0, ..., |\mathcal{N}|\}$ in period t for product p (see Figure 4.1). In a similar way, on the y-axis, which represents the values of the variance of the replenishment cycle demand ($\mathbf{m}_{tp}$), we define $|\mathcal{M}|$ predetermined approximation points m_{tpj} where $m_{tpj} \in [0, m_{tp}^{\max}]$ for every $j = \{0, ..., |\mathcal{M}|\}$ in period t for product p. We set $b_{tp}^{\max} = \sum_{i \in \mathcal{T}|i \geq t} d_{ip}$ and $m_{tp}^{\max} = \sum_{i \in \mathcal{T}|i \geq t} \sigma_{ip}^2$ for every product p in every period t.

For a given b_{tpi} and m_{tpj}, we use a numerical method to calculate the corresponding order-up-to-level y_{tpij}^{TL} (see the mathematical relation (4.23)) based on the given target service level and demand distribution.

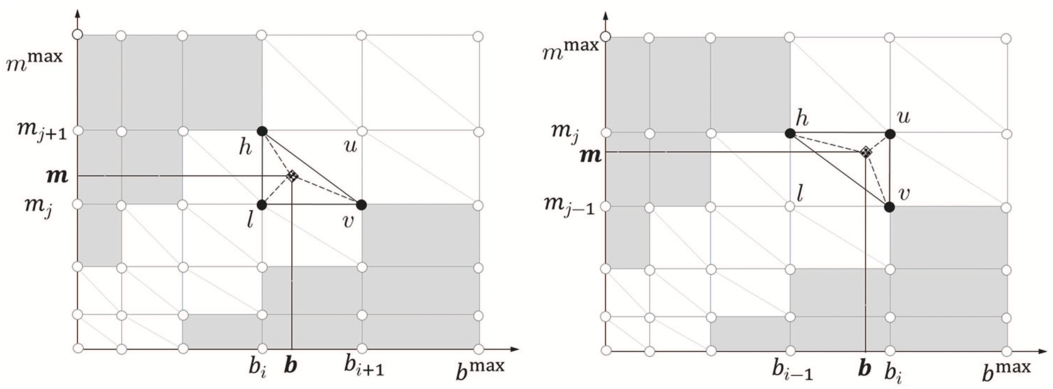


Figure 4.1: Approximation grid: Selecting a lower triangle (left-hand side), selecting an upper triangle (right-hand side). Indexes p and t have been dropped.

For every upper (lower) triangle, we define the binary decision variable as $\boldsymbol{\omega}_{ij}^{\mathrm{u}}$ ($\boldsymbol{\omega}_{ij}^{\mathrm{l}}$) and the three respective weights $\boldsymbol{\lambda}_{ij}^{\mathrm{u}}$, $\boldsymbol{\lambda}_{ij}^{\mathrm{uh}}$ and $\boldsymbol{\lambda}_{ij}^{\mathrm{uv}}$ ($\boldsymbol{\lambda}_{ij}^{\mathrm{l}}$, $\boldsymbol{\lambda}_{ij}^{\mathrm{lh}}$ and $\boldsymbol{\lambda}_{ij}^{\mathrm{lv}}$). For the sake of readability, we drop the indexes of p and t. On the left-hand side of Figure 4.1, we see a selected lower triangle ($\boldsymbol{\omega}_{ij}^{\mathrm{l}} = 1$), where the value of $\mathbf{b}$ is obtained by summing up the fraction of point b_i given by weights $\boldsymbol{\lambda}_{ij}^{\mathrm{l}}$ and $\boldsymbol{\lambda}_{ij}^{\mathrm{lv}}$, as well as the fraction of point b_{i+1} given by weight $\boldsymbol{\lambda}_{ij}^{\mathrm{lh}}$. The value of $\mathbf{m}$ in this triangle is obtained by summing up the fraction of point m_j given by weights $\boldsymbol{\lambda}_{ij}^{\mathrm{l}}$ and $\boldsymbol{\lambda}_{ij}^{\mathrm{lh}}$, as well as the fraction of point m_{j+1} given by weight $\boldsymbol{\lambda}_{ij}^{\mathrm{lv}}$. If the upper triangle is chosen, i.e., $\boldsymbol{\omega}_{ij}^{\mathrm{u}} = 1$ (on the right-hand side of Figure 4.1), then the weights of its vertices are determined in a similar way. Depending on which triangle (upper or lower) is chosen, the value of $\mathbf{y}^{\mathrm{TL}}$ is approximated by the convex combination of the respective vertices and the weights that are used for determining the corresponding values of $\mathbf{b}$ and $\mathbf{m}$.

We can improve the approximation grid by cutting off unnecessary small rectangles from the approximation grid (see the shaded rectangles in Figure 4.1). For example, think of an approximation point on one of the axes. An approximation point is a predetermined value with the boundary of the mean or variance of the replenishment cycle demand. Given the mean or variance per period, we can calculate the length of the replenishment cycle from the value of the approximation point. Having determined the length of the replenishment cycle for that approximation point, we can calculate the corresponding value on the other axis. All values (approximation points) beyond that value are irrelevant as far as the original approximation point is concerned, and can be deducted from the approximation grid. The final effective approximation grid will resemble a diagonal shape, as illustrated in Figure 4.1.

The following constraints present the linearisation of the bivariate non-linear function $\mathbf{y}^{\mathrm{TL}}_{tp}(\mathbf{b}_{tp}, \mathbf{m}_{tp})$. The corresponding notation is given in Table 4.6. Note that $|\mathcal{N}_{tp}|$ and $|\mathcal{M}_{tp}|$ indicate the total number of relevant approximation points for product p in period t on the x- and y-axis, respectively.

$$\forall p \in \mathcal{P}, t \in \mathcal{T}:$$

$$\sum_{i\in\mathcal{N}_{tp}} \sum_{j\in\mathcal{M}_{tp}} \boldsymbol{\omega}^{\mathrm{u}}_{tpij} + \sum_{i\in\bar{\mathcal{N}}_{tp}} \sum_{j\in\bar{\mathcal{M}}_{tp}} \boldsymbol{\omega}^{\mathrm{l}}_{tpij} = 1, \tag{4.41}$$

$$\boldsymbol{\lambda}^{\mathrm{u}}_{tpij} + \boldsymbol{\lambda}^{\mathrm{uv}}_{tpij} + \boldsymbol{\lambda}^{\mathrm{uh}}_{tpij} = \boldsymbol{\omega}^{\mathrm{u}}_{tpij}, \qquad i \in \mathcal{N}_{tp}, j \in \mathcal{M}_{tp} \tag{4.42}$$

$$\boldsymbol{\lambda}^{\mathrm{l}}_{tpij} + \boldsymbol{\lambda}^{\mathrm{lv}}_{tpij} + \boldsymbol{\lambda}^{\mathrm{lh}}_{tpij} = \boldsymbol{\omega}^{\mathrm{l}}_{tpij}, \qquad i \in \bar{\mathcal{N}}_{tp}, j \in \bar{\mathcal{M}}_{tp} \tag{4.43}$$

$$\mathbf{b}_{tp} = \sum_{i\in\mathcal{N}_{tp}} \sum_{j\in\mathcal{M}_{tp}} ((\boldsymbol{\lambda}^{\mathrm{u}}_{tpij} + \boldsymbol{\lambda}^{\mathrm{uv}}_{tpij}) \cdot b_{tpi} + \boldsymbol{\lambda}^{\mathrm{uh}}_{tpij} \cdot b_{tp,i-1}) +$$
$$\sum_{i\in\bar{\mathcal{N}}_{tp}} \sum_{j\in\bar{\mathcal{M}}_{tp}} ((\boldsymbol{\lambda}^{\mathrm{l}}_{tpij} + \boldsymbol{\lambda}^{\mathrm{lv}}_{tpij}) \cdot b_{tpi} + \boldsymbol{\lambda}^{\mathrm{lh}}_{tpij} \cdot b_{tp,i+1}), \tag{4.44}$$

$$\mathbf{m}_{tp} = \sum_{i\in\mathcal{N}_{tp}} \sum_{j\in\mathcal{M}_{tp}} ((\boldsymbol{\lambda}^{\mathrm{u}}_{tpij} + \boldsymbol{\lambda}^{\mathrm{uh}}_{tpij}) \cdot m_{tpj} + \boldsymbol{\lambda}^{\mathrm{uv}}_{tpij} \cdot m_{tp,j-1}) +$$
$$\sum_{i\in\bar{\mathcal{N}}_{tp}} \sum_{j\in\bar{\mathcal{M}}_{tp}} ((\boldsymbol{\lambda}^{\mathrm{l}}_{tpij} + \boldsymbol{\lambda}^{\mathrm{lh}}_{tpij}) \cdot m_{tpj} + \boldsymbol{\lambda}^{\mathrm{lv}}_{tpij} \cdot m_{tp,j+1}), \tag{4.45}$$

$$\mathbf{y}^{\mathrm{TL}}_{tp} = \sum_{i\in\mathcal{N}_{tp}} \sum_{j\in\mathcal{M}_{tp}} (\boldsymbol{\lambda}^{\mathrm{u}}_{tpij} y^{\mathrm{TL}}_{pij} + \boldsymbol{\lambda}^{\mathrm{uh}}_{tpij} y^{\mathrm{TL}}_{p,i-1,j} + \boldsymbol{\lambda}^{\mathrm{uv}}_{tpij} y^{\mathrm{TL}}_{pi,j-1}) +$$
$$\sum_{i\in\bar{\mathcal{N}}_{tp}} \sum_{j\in\bar{\mathcal{M}}_{tp}} (\boldsymbol{\lambda}^{\mathrm{l}}_{tpij} y^{\mathrm{TL}}_{pij} + \boldsymbol{\lambda}^{\mathrm{lh}}_{tpij} y^{\mathrm{TL}}_{p,i+1,j} + \boldsymbol{\lambda}^{\mathrm{lv}}_{tpij} y^{\mathrm{TL}}_{pi,j+1}), \tag{4.46}$$

$$\mathbf{b}_{tp}, \mathbf{m}_{tp}, \mathbf{y}^{\mathrm{TL}}_{tp} \geq 0, \qquad (4.47)$$

$$\boldsymbol{\omega}^{\mathrm{u}}_{tpij} \in \{0,1\}, \boldsymbol{\lambda}^{\mathrm{u}}_{tpij}, \boldsymbol{\lambda}^{\mathrm{uh}}_{tpij}, \boldsymbol{\lambda}^{\mathrm{uv}}_{tpij} \geq 0, \qquad i \in \mathcal{N}_{tp}, j \in \mathcal{M}_{tp} \qquad (4.48)$$

$$\boldsymbol{\omega}^{\mathrm{l}}_{tpij} \in \{0,1\}, \boldsymbol{\lambda}^{\mathrm{l}}_{tpij}, \boldsymbol{\lambda}^{lh}_{tpij}, \boldsymbol{\lambda}^{\mathrm{lv}}_{tpij} \geq 0. \qquad i \in \bar{\mathcal{N}}_{tp}, j \in \bar{\mathcal{M}}_{tp} \qquad (4.49)$$

Constraints (4.41) ensure that only one triangle on the entire approximation grid is selected. Constraints (4.42) and (4.43) guarantee that when a triangle is chosen, the corresponding weights are added up to one and no values for the weights of a triangle are allowed if that triangle has not been chosen. Constraints (4.44) and (4.45) determine the corresponding weights used for obtaining the values of $\mathbf{b}_{tp}$ and $\mathbf{m}_{tp}$ respectively. Constraints (4.46) determine the approximated value for $\mathbf{y}^{\mathrm{TL}}_{tp}$ in period t for every product p. Non-negativity and binary variables are given in (4.47) to (4.49).

Table 4.6: Notation used for the approximation grid.

Sets:	
$i \in \mathcal{N}_{tp} = \{1, ..., \lvert\mathcal{N}_{tp}\rvert\}$	approximation points on x-axis for product p in period t (upper triangles)
$j \in \mathcal{M}_{tp} = \{1, ..., \lvert\mathcal{M}_{tp}\rvert\}$	approximation points on y-axis for product p in period t (upper triangles)
$i \in \bar{\mathcal{N}}_{tp} = \{0, ..., \lvert\mathcal{N}_{tp}\rvert - 1\}$	approximation points on x-axis for product p in period t (lower triangles)
$j \in \bar{\mathcal{M}}_{tp} = \{0, ..., \lvert\mathcal{M}_{tp}\rvert - 1\}$	approximation points on y-axis for product p in period t (lower triangles)
Parameters:	
b_{tpi}	predetermined approximation point i on the x-axis
m_{tpj}	predetermined approximation point j on the y-axis
y^{TL}_{tpij}	predetermined order-up-to level for the combination of b_{tpi} and m_{tpj}
Decision variables:	
$\boldsymbol{\lambda}^{\mathrm{u}}_{tpij}, \boldsymbol{\lambda}^{\mathrm{uv}}_{tpij}, \boldsymbol{\lambda}^{\mathrm{uh}}_{tpij}$	weights associated with the vertices of the upper triangles
$\boldsymbol{\lambda}^{\mathrm{l}}_{tpij}, \boldsymbol{\lambda}^{\mathrm{lv}}_{tpij}, \boldsymbol{\lambda}^{\mathrm{lh}}_{tpij}$	weights associated with the vertices of the lower triangles
$\boldsymbol{\omega}^{\mathrm{u}}_{tpij}$	binary decision variable equals one if the approximation point with indexes of i and j for product p in period t is in an upper triangle, otherwise zero.

$\boldsymbol{\omega}^{\mathrm{l}}_{tpij}$	binary decision variable equals one if the approximation point with indexes of i and j for product p in period t is in a lower triangle, otherwise zero.

4.4.3 The Integrated Model with Re-planning-Opportunity Adjustment

The MILP we presented in Section 4.4.2 for the integrated model does not take the future re-planning opportunities under rolling horizon planning into account. The re-planning opportunities give the flexibility to change previously planned schedules according to the actual demand realisation. In other words, the re-planning opportunities can protect against some demand uncertainty if sufficient capacity is available under rolling horizon planning. In such cases, we need less safety stock. In order to reduce safety stock to account for the flexibility, we reduce the target inventory level, $\mathbf{y}^{\mathrm{TL}}_{tp}(\mathbf{b}_{tp}, \mathbf{m}_{tp})$, in the integrated model by discounting the values of $\mathbf{b}_{tp}$ and $\mathbf{m}_{tp}$ as follows.

$$\mathbf{y}^{\alpha\mathrm{TL}}_{tp} = \mathbf{y}^{\mathrm{TL}}_{tp}(\alpha_p \cdot \mathbf{b}_{tp}, \alpha_p \cdot \mathbf{m}_{tp}), \qquad \forall t \in \mathcal{T}, p \in \mathcal{P} \tag{4.50}$$

where $\mathbf{y}^{\alpha\mathrm{TL}}_{tp}$ is the adjusted order-up-to level in period t for product p and $(0 < \alpha_p \leq 1)$ is an exogenous product-specific re-planning opportunity coefficient. In what follows, we introduce a heuristic method for approximating the optimal values of α_p under rolling horizon planning.

Re-planning opportunity coefficient (α_p)

If the capacity is not binding, the average length of the replenishment cycles of product p over the planning horizon can be approximated by the average time between orders (L_p). In our case, L_p can be easily calculated based on the total demand and the minimum production quantity. As illustrated in Figure (4.2), the demand uncertainty during L_p is not only covered by the safety stock, but also by the safety capacity with respect to re-planning opportunities. The re-planning opportunity coefficient α_p anticipates the fraction of L_p covered by the safety capacity, i.e., $(1 - \alpha_p) \cdot L_p$. The demand uncertainty of the remaining part, i.e., $\alpha_p \cdot L_p$, is covered by the dynamic safety stock.

If the capacity is limited, we cannot reduce the value of α_p beyond a certain point. In this case, we first approximate an average excess capacity level (the safety capacity). Let

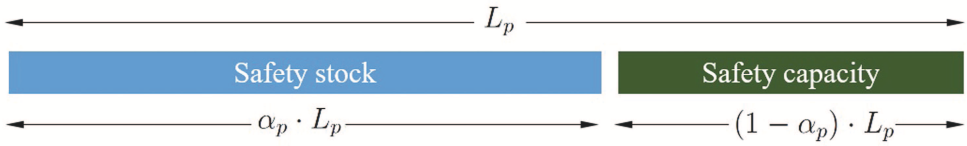


Figure 4.2: Replenishment cycle of product p with an average length of L_p.

$\bar{K}_p$ denote the safety capacity available during L_p for product p. The following equation approximates $\bar{K}_p$ by multiplying the length of L_p with the average excess capacity per product and period. The average excess capacity per product and period is determined by uniformly distributing the excess capacity for the entire planning horizon (the numerator in (4.51)) over all products and periods (the denominator).

$$\bar{K}_p = L_p \cdot \frac{\sum_{t\in T} K_t - \sum_{t\in\mathcal{T}} \sum_{p\in\mathcal{P}} k_P \cdot d_{tp} - \sum_{p\in\mathcal{P}} z_p \cdot \frac{|\mathcal{T}|}{L_p}}{|\mathcal{T}| \cdot |\mathcal{P}|}. \qquad \forall p \in \mathcal{P} \qquad (4.51)$$

We ensure that $\bar{K}_p$ is sufficient for producing the demand during $(1-\alpha_p) \cdot L_p$.

$$(1-\alpha_p) \cdot L_p \cdot \bar{d}_p \cdot k_p \leq \bar{K}_p. \qquad \forall p \in \mathcal{P} \qquad (4.52)$$

$$\alpha_p \geq \frac{L_p \cdot \bar{d}_p \cdot k_p - \bar{K}_p}{L_p}. \qquad \forall p \in \mathcal{P} \qquad (4.53)$$

If the safety capacity is negligible, α_p becomes almost one, which means that there is a small degree of flexibility that can be used for influencing the replenishment cycles under rolling horizon planning. However, if the capacity is unlimited, α_p is not restricted by (4.53). This implies that the safety capacity alone can cover the demand uncertainty. This might not always be true since we may need more setups than optimally required. Since there is a minimum production quantity, an extra setup can result in an excess inventory at the end of the planning horizon. Thus, in the following, we look for a trade-off between the inventory cost of increasing the number of setups and the benefits of decreasing the dynamic safety stock in order to find a lower boundary on α_p. Let n_p^{setup} denote the increased number of setups for product p over the planning horizon. Mathematically,

$$n_p^{\text{setup}} = \lceil \frac{|\mathcal{T}|}{\alpha_p \cdot L_p} - \frac{|\mathcal{T}|}{L_p} \rceil, \qquad \forall p \in \mathcal{P} \qquad (4.54)$$

where $\lceil x \rceil$ indicates the smallest integer value that is higher than or equal to x. Note that $n_p^{\text{setup}} \leq |\mathcal{T}|$, which implies that $\alpha_p \geq \frac{1}{1+L_p}$.

The following equation (4.55) approximates the total cost incurred by incorporating α_p. The first term indicates the approximation of the increased inventory holding cost brought about by the increased number of setups in relation to $q_p^{\min}$. The second term indicates the savings on the inventory costs through a reduced dynamic safety stock.

$$\mathrm{TC}(\alpha_p) = \frac{1}{2} \cdot n_p^{\mathrm{setup}} \cdot q_p^{\min} \cdot c_p^{\mathrm{h}} - [y_p^{\mathrm{sst,L}} - y_p^{\mathrm{sst},\alpha\mathrm{L}}] \cdot |\mathcal{T}| \cdot c_p^h, \qquad \forall p \in \mathcal{P} \tag{4.55}$$

where, $y_p^{\mathrm{sst,L}}$ and $y_p^{\mathrm{sst},\alpha\mathrm{L}}$ denote the dynamic safety stocks necessary to cover demand uncertainty of product p during L_p and $\alpha_p \cdot L_p$, respectively, in relation to a target service level.

We use a numerical search to approximate the value $\alpha_p^* \in \Big[max\{\frac{1}{1+L_p}, \frac{L_p \cdot \bar{d}_p \cdot k_p - \bar{K}_p}{L_p}\}, 1\Big]$ that returns the lowest value for the total cost function in (4.55). The second lower boundary is a logical consequence after applying the constraints (4.53).

In the following, we present the full MILP for the integrated model with re-planning opportunity adjustment. Note that if we set $\alpha_p = 1$, then the the integrated model with re-planning opportunity adjustment resembles the standard integrated model.

The MILP for the integrated model with re-planning opportunity adjustment

The objective function is given in (4.56). The base constraints are shown in (4.57) to (4.61). The relevant constraints to the service-level are given in (4.62), (4.63) and (4.64). The mean of the replenishment cycle demand is determined by means of (4.65) to (4.68). The variance of the replenishment cycle demand is determined by means of (4.69) to (4.72). Linearisation of $\mathbf{y}_{tp}^{\alpha\mathrm{TL}}$ is similar to the model formulation given in (4.41) to (4.49) in the previous section on pages 55 and 56. Non-negativity constraints and binary decision variables are presented in (4.73) to (4.76).

$$\min\ TC =$$
$$\sum_{p\in\mathcal{P}} \sum_{t\in\mathcal{T}} \left(c_p^{\mathrm{h}} \cdot \mathbf{y}_{tp}^{+} + c_p^{\mathrm{minq}} \cdot \boldsymbol{\zeta}_{tp} + c_p^{\mathrm{bl}} \cdot \boldsymbol{\psi}_{tp}^{\mathrm{bl}} + c_p^{\mathrm{ds}} \cdot \boldsymbol{\psi}_{tp}^{\mathrm{ds}} \right) + \sum_{p\in\mathcal{P}} c_p^{y^{\mathrm{fi}}} \cdot \boldsymbol{\rho}_p, \tag{4.56}$$

subject to

$$\mathbf{y}^+_{tp} - \mathbf{y}^-_{tp} = y_{0p} + \sum_{i\in\mathcal{T}|i\leq t} (\mathbf{q}_{ip} - d_{ip}), \qquad \forall t \in \mathcal{T}, p \in \mathcal{P} \tag{4.57}$$

$$\sum_{p\in\mathcal{P}} (k_p \cdot \mathbf{q}_{tp} + z_p \cdot \boldsymbol{\delta}_{tp}) \leq K_t, \qquad \forall t \in \mathcal{T} \tag{4.58}$$

$$k_p \cdot \mathbf{q}_{tp} \leq K_t \cdot \boldsymbol{\delta}_{tp}, \qquad \forall t \in \mathcal{T}, p \in \mathcal{P} \tag{4.59}$$

$$\mathbf{q}_{tp} \geq \boldsymbol{\delta}_{tp} \cdot q^{\min}_p - \boldsymbol{\zeta}_{tp}, \qquad \forall t \in \mathcal{T}, p \in \mathcal{P} \tag{4.60}$$

$$\mathbf{y}^+_{|\mathcal{T}|,p} - \mathbf{y}^-_{|\mathcal{T}|,p} \geq y^{\text{fi}}_p - \rho_p, \qquad \forall p \in \mathcal{P} \tag{4.61}$$

$$y_{0p} + \mathbf{q}_{tp} \geq \mathbf{y}^{\alpha\text{TL}}_{tp} + (1-\alpha_p)\cdot d_{tp} - \boldsymbol{\psi}^{\text{bl}}_{tp} - \boldsymbol{\psi}^{\text{ds}}_{tp}, \qquad \forall p \in \mathcal{P}, t = 1 \tag{4.62}$$

$$\begin{aligned} & y_{0p} + \sum_{i\in\mathcal{T}|i\leq t} \mathbf{q}_{ip} - \sum_{i\in\mathcal{T}|i<t} d_{ip} \geq \\ & \mathbf{y}^{\alpha\text{TL}}_{tp} + (1-\alpha_p)\cdot d_{tp} - \boldsymbol{\psi}^{\text{bl}}_{tp} - \boldsymbol{\psi}^{\text{ds}}_{tp} - (1-\boldsymbol{\delta}_{tp})\cdot M, \qquad \forall t \in \mathcal{T}, p \in \mathcal{P} | t > 1 \end{aligned} \tag{4.63}$$

$$\boldsymbol{\psi}^{\text{ds}}_{tp} \leq \mathbf{y}^{\text{ds}}_{tp}, \qquad \forall t \in \mathcal{T},\ p \in \mathcal{P} \tag{4.64}$$

$$\mathbf{u}_{tp} \leq \boldsymbol{\delta}_{tp}, \qquad \forall t \in \mathcal{T}, p \in \mathcal{P} \tag{4.65}$$

$$\mathbf{u}_{tp} \geq \bar{\mathbf{b}}_{tp} - (1-\boldsymbol{\delta}_{tp}), \qquad \forall t \in \mathcal{T}, p \in \mathcal{P} \tag{4.66}$$

$$\mathbf{u}_{tp} \leq \bar{\mathbf{b}}_{tp} + (1-\boldsymbol{\delta}_{tp}), \qquad \forall t \in \mathcal{T}, p \in \mathcal{P} \tag{4.67}$$

$$\bar{\mathbf{b}}_{tp} \cdot b^{\max}_{tp} = \sum_{i\in\mathcal{T}|i\geq t} d_{ip} - \sum_{i\in\mathcal{T}|i>t} \mathbf{u}_{ip} \cdot b^{\max}_{ip}, \qquad \forall t \in \mathcal{T}, p \in \mathcal{P} \tag{4.68}$$

$$\mathbf{c}_{tp} \leq \boldsymbol{\delta}_{tp}, \qquad \forall t \in \mathcal{T}, p \in \mathcal{P} \tag{4.69}$$

$$\mathbf{c}_{tp} \geq \bar{\mathbf{m}}_{tp} - (1-\boldsymbol{\delta}_{tp}), \qquad \forall t \in \mathcal{T}, p \in \mathcal{P} \tag{4.70}$$

$$\mathbf{c}_{tp} \leq \bar{\mathbf{m}}_{tp} + (1-\boldsymbol{\delta}_{tp}), \qquad \forall t \in \mathcal{T}, p \in \mathcal{P} \tag{4.71}$$

$$\bar{\mathbf{m}}_{tp} \cdot m^{\max}_{tp} = \sum_{i\in\mathcal{T}|i\geq t} \sigma^2_{ip} - \sum_{i\in\mathcal{T}|i>t} \mathbf{c}_{ip} \cdot m^{\max}_{ip}, \qquad \forall t \in \mathcal{T}, p \in \mathcal{P} \tag{4.72}$$

$$\mathbf{y}^+_{tp} \geq 0, \mathbf{y}^-_{tp} \geq 0, \mathbf{q}_{tp} \geq 0, \boldsymbol{\delta}_{tp} \in \{0,1\}, \qquad \forall t \in \mathcal{T}, p \in \mathcal{P} \tag{4.73}$$

$$\boldsymbol{\zeta}_{tp} \geq 0, \boldsymbol{\psi}^{\text{ds}}_{tp} \geq 0, \boldsymbol{\psi}^{\text{bl}}_{tp} \geq 0, \qquad \forall t \in \mathcal{T}, p \in \mathcal{P} \tag{4.74}$$

$$\mathbf{y}^{\text{TL}}_{tp} \geq 0, \mathbf{u}_{tp} \geq 0, \bar{\mathbf{b}}_{tp} \geq 0, \mathbf{c}_{tp} \geq 0, \bar{\mathbf{m}}_{tp} \geq 0, \qquad \forall t \in \mathcal{T}, p \in \mathcal{P} \tag{4.75}$$

$$\boldsymbol{\rho}_p \geq 0. \qquad \forall p \in \mathcal{P} \tag{4.76}$$

4.5 Numerical Study

4.5.1 Experimental Design

We use the illustration in Figure 4.3 to explain the rolling horizon setting. Let n refer to the n^{th} rolling horizon schedule. In the first schedule ($n = 1$), we solve the problem for a planning horizon with $|\mathcal{T}| = 12$ periods. We implement the production plan for the first period, i.e., the frozen horizon. Note that the production plan during the frozen horizon is no longer updated. We roll forward the planning horizon for one period (re-planning horizon) in order to generate the second rolling horizon schedule ($n = 2$). The re-planning horizon is the time that elapses between the starting points of two consecutive rolling horizon schedules. Then, we implement the production plan of its frozen horizon, roll forward and repeat until we implement the production plans of all periods of the evaluation interval. If we define an evaluation interval of $|\mathcal{E}| = 50$ periods, we generate in total of 50 rolling schedules $n = \{1, 2, ..., 50\}$.

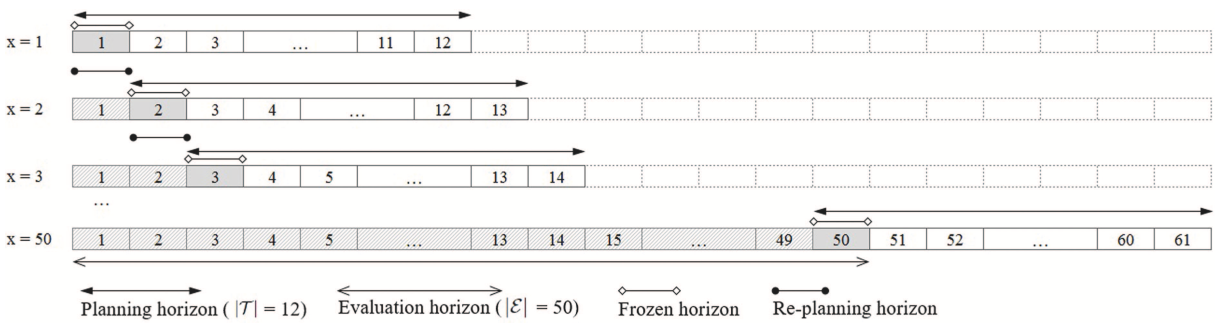


Figure 4.3: Rolling horizon setting.

Besides the rolling horizon setting, we further define an open loop setting, where we set both the frozen horizon and the re-planning horizon to 10 periods. The open loop setting can be seen as a variant of the static uncertainty strategy introduced by Bookbinder and Tan (1988), where the length of the frozen horizon is equal to the length of the planning horizon. However, in the open loop setting, we take ten periods instead of twelve periods, as this reduces the influence of the final inventory level constraints on the starting inventory level of the next schedule. Table 4.7 summarises the experimental design.

We use the following procedure to generate the forecast and demand values. We assume that forecast values are not updated over time and that the length of the planning horizon ($|\mathcal{T}| = 12$) remains constant. Thus, we need $|\mathcal{S}| = 61$ values of the forecasted demand $\tilde{d}_{jp}$

Table 4.7: Experimental design.

Initialisation of Demand Parameters:		
Mean demand	$\bar{d}_p$	$= 30$
Inter-period demand variation	CV^{ip}	$= 0.3$
Coefficient of variation	CV^{d}	$= 0.3$
Levels of Main Input Factors:		
Time between orders	L	$\in \{1, 2, 3, 6\}$
Capacity level	Capa.	$\in$ {uncapacitated, capacitated}
Target service level	β^{c}	$\in \{0.90, 0.98\}$
Initialisation of Basic Parameters:		
Inventory holding cost	c_p^{h}	1
Production capacity utilisation	k_p	1
Setup time	z_p	$0.15 \cdot K_t$
Min. production quantity	q_p^{min}	$w^{\text{minq}} \cdot \bar{d}_p$
Min. production quantity coefficient	w^{minq}	{0,2,3,6}
Backlog/backorder penalty cost	c_p^{bl}	120
Shortfall from exogenous safety stock	c_p^{sst}	$\{0.1, 0.2, 0.3\} \cdot c_p^{bl}$
Shortfall from dynamic safety stock	c_p^{ds}	c_p^{sst}
Shortfall from min. production quantity	c_p^{minq}	600
Shortfall from final inventory	$c_p^{y^{\text{fi}}}$	c_p^{minq}

$\forall p \in \mathcal{P}, j \in \mathcal{S} = \{1, ..., |\mathcal{S}|\}$. Note that the last run under the rolling horizon takes place between periods 50 and 61 (see Figure 4.3). Let CV^{ip} denote the coefficient variation of the inter-period demands. In order to generate dynamic forecast values, we use Normal Distribution (ND) with a mean demand indicated by $\bar{d}_p$, and a standard deviation of $\bar{d}_p \cdot CV^{\text{ip}}$.

$$\tilde{d}_{jp} = ND(\bar{d}_p, \bar{d}_p \cdot CV^{\text{ip}}). \qquad \forall p \in \mathcal{P}, j \in \mathcal{S} \tag{4.77}$$

Since the forecast is not updated, the expected demand values d_{tp} $\forall p \in \mathcal{P}$, $t \in \mathcal{T}$ for a given schedule n are taken from $\tilde{d}_{jp}$ as follows.

$$d_{tp} = \tilde{d}_{t+n-1,p}, \qquad \forall p \in \mathcal{P}, t \in \mathcal{T} \tag{4.78}$$

where $n \in \{1, 2, .., 50\}$ under the rolling horizon and $n \in \{1, 2, .., 5\}$ under the open loop.

We assume that the actual demand $\hat{d}_{ip}$ for every $p \in \mathcal{P}$ and $i \in \mathcal{E} = \{1, ..., |\mathcal{E}|\}$ follows a Gamma Distribution (GD) with mean $\tilde{d}_{ip}$ and standard deviation $\tilde{d}_{ip} \cdot CV^{\text{d}}$, where CV^{d}

indicates the coefficient of variation of demand that relates to the randomness of the actual demand.

$$\hat{d}_{ip} = GD(\tilde{d}_{ip}, \tilde{d}_{ip} \cdot CV^{\mathrm{d}}). \qquad \forall p \in \mathcal{P}, i \in \mathcal{E} \tag{4.79}$$

We randomly generate 20 independent replications of the actual demand series for each combination of the main input factors, which is a high enough number of replications to derive consistent results. The main input factors are defined as follows.

a. *Time between orders (L)*

We use q_p^{min} to define the time between orders. q_p^{min} is calculated as the product of the coefficient w^{minq} and the mean demand. Thus, we set $w^{\mathrm{minq}} = 0$ to obtain L =1. We set $w^{\mathrm{minq}} = \{2, 3, 6\}$ to obtain $L = \{2, 3, 6\}$.

b. *Capacity level*

We take two cases of the uncapacitated and capacitated levels into account. In a capacitated case, we determine the available capacity level as follows. We use w^{util} as an indicator, which is defined as a fraction of the production capacity requirement for all products based on their average demand per period as follows.

$$K_t = C = \frac{\sum_{j \in \mathcal{S}} \sum_{p \in \mathcal{P}} \tilde{d}_{jp}}{|\mathcal{S}|} \cdot w^{\mathrm{util}}. \qquad \forall t \in \mathcal{T} \tag{4.80}$$

To ensure a sufficient capacity for producing q_p^{min} within a period, we set the values of w^{util} based on the values of w^{minq}, which is done as follows. If $w^{\mathrm{minq}} = 0$, we set $w^{\mathrm{util}} = 1.3$. If $w^{\mathrm{minq}} \in \{2, 3, 6\}$, we set $w^{\mathrm{util}} = w^{\mathrm{minq}} + 0.3$. This means that the capacity level in every period is the total capacity needed to produce q_p^{min} plus 30% of the capacity requirement for the mean demand. Moreover, as we increase the capacity level according to w^{minq}, we also increase the number of products in order to keep the problem capacitated. Thus, if we investigate an uncapacitated case or a capacitated case with $w^{\mathrm{minq}} = 0$, we take only a single product ($|\mathcal{P}| = 1$) into account. If $w^{\mathrm{minq}} = 2$, we set $|\mathcal{P}| = 2$. Finally, if we use $w^{\mathrm{minq}} = 3$ or 6, we take $|\mathcal{P}| = 3$.

c. *Target cycle fill-rate (β^c-service level)*

We use β^c to determine order-up-to-levels for the integrated models. We use a simple rule-of-thumb approach, i.e., a common day of supply (w^{cDoS}), to calculate exogenous safety stocks for the sequential approach. w^{cDoS} is defined as a factor multiplied with the length of one period. The exogenous safety stock in every period is calculated by multiplying w^{cDoS} with the forecasted demand of the period.

The holding cost of product $p \in \mathcal{P}$ per period and the capacity consumption per unit of product p are set to one. The setup time of every product p is set to 15% of the available capacity level in period t. We assume that setup costs are zero. The backlog penalty cost per unit of product p and per period is defined as the total holding cost of product p over the entire planning horizon multiplied by ten ($c_p^{\text{bl}} = 120$). The initial inventory level of product p at the beginning of the evaluation interval is set to the sum of the expected demand over the first two periods ($y_{0p} = \tilde{d}_{p1} + \tilde{d}_{p2}$). The final inventory level for the open loop setting is set as follows.

$$y_p^{\text{fi}} = \frac{\sum_{t \in \mathcal{T}} \tilde{d}_{tp}}{|\mathcal{T}|}. \qquad \forall p \in \mathcal{P} \tag{4.81}$$

The shortfalls of $q_p^{\min}$ and of y_p^{fi} are penalised with five times the value of the backlog penalty cost ($c_p^{\text{minq}} = c_p^{y^{\text{fi}}} = 600$). The penalty cost to be paid for a shortfall of the exogenous safety stock level (or the dynamic safety stock level in the order-up-to-level) is set to $c_p^{\text{sst}} = c_p^{\text{ds}} =$0.1, 0.2, and 0.3 of the backlog penalty cost.

We report two main performance measures as the outcomes of the evaluation setting: The Total Cost (TC) and the average realised cycle fill-rate (β^{RL}) over the evaluation interval. The TC is defined as the cost of holding the actual inventory on-hand level at the end of the periods within the evaluation interval. β^{RL} is defined as the average of realised cycle fill-rates over all replenishment cycles, over all products and over the evaluation interval. If n^{RC} indicates the total number of the realised replenishment cycles over the evaluation interval, we can mathematically write

$$\beta^{\text{RL}} = \frac{1}{|\mathcal{P}| \cdot n^{\text{RC}}} \cdot \sum_{p \in \mathcal{P}} \sum_{i \in \{1,2,\ldots,n^{\text{RC}}\}} (1 - \frac{\varphi_{ip}^{\text{RC}}}{d_{ip}^{\text{RC}}}), \tag{4.82}$$

where φ_{ip}^{RC} and d_{ip}^{RC} denote the total backorder quantities and the total actual demand of product p during the realised replenishment cycle i.

The comparison between two selected modelling approaches, i.e., the sequential approach and the integrated model, is done under consideration of the total cost difference (ΔTC) when both approaches return an identical β^{RL}. Specifically, ΔTC is the TC obtained by the sequential approach minus the TC from the integrated model divided by the TC from the integrated model, given in percentage. If this value is positive, then the integrated model has a lower TC than the sequential approach. If ΔTC is negative, then the sequential approach outperforms the integrated model.

4.5.2 Numerical Results

Computational performance of the integrated model

We use the MIP solver FICO Xpress Optimizer 64-Bit v.28.01.04 to solve all the models on a platform with sufficient RAM and the following CPU specification: Intel Core i7-4770 CPU @ 3.40 GHz 3.40 GHz, 64-bit. The maximum runtime is set to 3,600 seconds.

Table 4.8 reports the averages and standard deviations of the runtime (in seconds) required for solving the MILP of the integrated model. The table also reports the average and standard deviations of the optimality gap (in %) and the average number of rows (constraints) and columns (variables) of the MILP.

The results are based on 50 runs under rolling horizon planning for the capacitated and uncapacitated cases and for three different values of $L = \{1, 2, 3\}$ with $\beta^c = 0.98$. We also distinguish between the General Approximation Grid (GAG) and the Improved Approximation Grid (IAG) introduced in Section 4.4.2.

Taking the results from Table 4.8 into account, we observe that the MILP is computationally efficient if we use the IAG. With the exception of the uncapacitated case with L=3, where an average optimality gap of 2.2% is returned, all instances are solved to optimality. Compared to the GAG, using the IAG can significantly increase the computational performance. The IAG reduces the required average numbers of rows and columns by half. This is due to the fact that the IAG cuts off the unnecessary binary decision variables on the approximation grid.

Soft service level constraints versus hard service level constraints in the integrated model

We want to find out if we need to use the soft service level constraints instead of the hard service level constraints in order to avoid infeasibility in case of limited capacity. We test

Table 4.8: Computational performance (runtime in seconds, gap in %).

		Capacitated				Uncapacitated			
		Runtime	gap	Rows	Columns	Runtime	gap	Rows	Columns
	L	avg [stDev]	avg [stDev]	avg	avg	avg [stDev]	avg [stDev]	avg	avg
GAG	1	8 [13]	0 [0]	2606	9745	10 [7]	0 [0]	2600	9745
	2	304 [573]	0 [0]	5196	19490	531 [983]	0.4 [1.5]	5182	19454
	3	2926[1062]	29.8[29.4]	8027	28634	3603[2]	42.1[10.4]	7883	27550
IAG	1	2 [1]	0 [0]	1435	5061	3 [1]	0 [0]	1429	5061
	2	25 [35]	0 [0]	2851	10111	19 [13]	0 [0]	2843	10111
	3	420 [411]	0 [0]	4291	15209	1463 [1417]	2.2 [6.4]	4334	15132

100 runs (new production schedules) under the open loop and 1,000 runs under rolling horizon planning for each β^c and L combination. We start by solving each run with the Hard Service Level Constraint (HSLC). If there is an infeasibility during the procedure, then we introduce the Soft Service Level Constraint (SSLC).

Table 4.9 presents the percentage of the total instances where the SSLC, rather than the HSLC, are needed. Furthermore, Table 4.9 shows the average capacity utilisation over the evaluation interval and over the independent replications of the actual demand series. Based on the results presented in Table 4.9, we make the following main observations. With the exception of L=6, where the average capacity utilisation is low (less than 50%), the SSLC is always needed in order to obtain feasible solutions. We notice that, in general, the necessity for the SSLC is reduced as the value of L increases. Moving from $\beta^c = 0.90$ to $\beta^c = 0.98$, we observe that the necessity for the SSLC is slightly higher if L=1 and lower if $L > 1$. These observations can be explained as follows: An increase of L or a move from $\beta^c = 0.90$ to $\beta^c = 0.98$ can increase inventory levels, the consequence of which is the avoidance of a shortfall of the order-up-to-level.

Table 4.9: Soft service level constraints versus hard service level constraints.

	β^c=0.90				β^c=0.98			
	open loop		rolling horizon		open loop		rolling horizon	
L	SSLC in %	Cap. util. avg. in %	SSLC in %	Cap. util. avg. in %	SSLC in %	Cap. util. avg. in %	SSLC in %	Cap. util. avg. in %
1	25.0	86.1	48.2	85.5	29.0	86.5	49.7	86.1
2	30.0	79.3	27.6	90.2	22.0	91.4	8.9	91.4
3	4.0	93.1	1.9	92.6	1.0	94.0	0.4	92.4
6	-	48.5	-	47.4	-	49.4	-	47.8

Realised service levels versus target service levels in the integrated model

In order to facilitate the comparison between the values of β^{RL} and β^c, we first analyse the evolution of the inventory level over the evaluation interval. Figure 4.4 shows the evolution of the average inventory level during the evaluation interval over the independent replications of the actual demand series. We assume $\beta^c = 0.90$ and L=1 for both the uncapacitated and the capacitated case under the open loop and the rolling horizon.

Taking the illustrations from Figure 4.4 into account, we make the following main observations. If the capacity is not binding, then rolling horizon planning leads to considerably lower average inventory and backlog levels than open loop planning. If the capacity becomes binding under the open loop and the rolling horizon, both the inventory and backlog level are increased. Under the rolling horizon, this implies that the capacity limitation significantly restricts flexibility when it comes to rescheduling. Therefore, the time backorders spend in the system until they are fulfilled are longer.

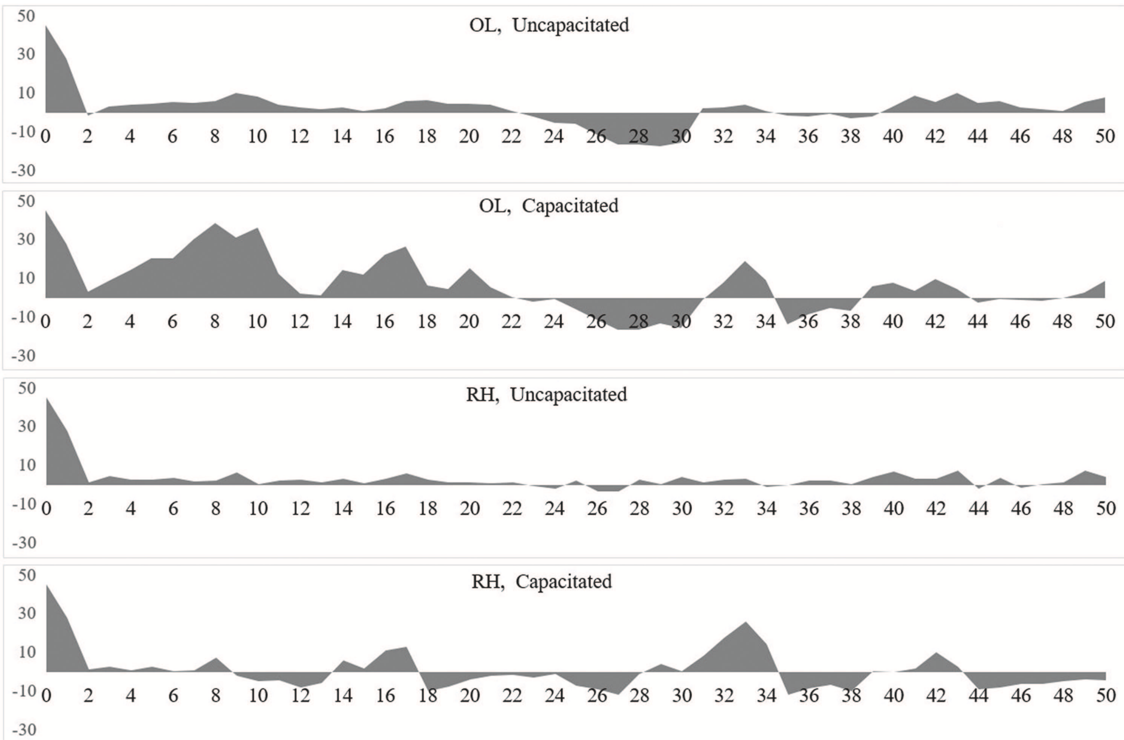


Figure 4.4: Evolution of the average inventory level ($\beta^c = 0.90$).

Table 4.10 presents the values of β^{RL} and the corresponding standard deviations over the evaluation interval and the independent replications of the actual demand series. The results are evaluated for each β^c, L and capacity level combination under both the rolling horizon and the open loop. Note that Table 4.9 shows the corresponding average capacity utilisation for the settings under the capacitated case.

Taking the open loop planning into account, we make the following main observations. β^{RL} usually overachieves β^c with $\beta^c = 0.90$. If $\beta^c = 0.98$, β^{RL} moves closer to the values of β^c and a considerable over-achievement is observed if $L = 6$. This observation can be explained by the fact that larger lot-sizes can prevent backorder quantities in the initial periods of the replenishment cycle, which leads to an improvement of β^{RL} values. If the capacity becomes binding (Capa.), the values of β^{RL} usually improve slightly. These improvements are the result of the cumulated inventories in periods with a low demand and can be applied in order to protect against backorders in periods with a peak demand.

Taking the rolling horizon planning into account, we make the following main observations. In every case, β^{RL} is higher than β^c. In contrast to open loop planning, the outstanding backorder quantities from the previous cycle are immediately fulfilled. Furthermore, we observe that an increase of L leads to very high values of β^{RL}. This implies that, under the rolling horizon, larger batch sizes, along with re-planning opportunities, can cover most of the demand uncertainty, even if β^c is set low. If the capacity becomes binding, β^{RL} decreases. In this case, it takes a long time for the outstanding backorder quantities to be fulfilled, which is due to restricted re-planning opportunities.

Table 4.10: Realised service level (β^{RL}) in %.

	β^c=0.90				β^c=0.98			
	open loop		rolling horizon		open loop		rolling horizon	
	Uncapa.	Capa.	Uncapa.	Capa.	Uncapa.	Capa.	Uncapa.	Capa.
L	avg. [st.Dev.]	avg. [st.Dev.]	avg. [st.Dev.]	avg. [st.Dev.]	avg. [st.Dev.]	avg. [st.Dev.]	avg. [st.Dev.]	avg. [st.Dev.]
1	95.1 [0.9]	95.8 [1.2]	94.8 [0.9]	94.9 [0.7]	96.3 [1.2]	96.9 [1.1]	98.9 [0.5]	98.1 [1.0]
2	91.1 [2.7]	89.8 [3.1]	99.0 [0.6]	96.7 [2.7]	97.0 [2.1]	96.5 [3.1]	99.9 [0.2]	99.7 [0.5]
3	91.5 [2.5]	95.9 [2.1]	99.4 [0.6]	98.7 [1.3]	98.3 [1.4]	98.8 [1.1]	100 [0.1]	99.8 [0.3]
6	95.1 [1.5]	96.6 [0.9]	99.6 [0.6]	99.6 [0.5]	99.3 [0.8]	99.6 [0.5]	99.9 [0.4]	99.7 [0.3]

Dynamic safety stocks versus exogenous safety stocks

We investigate what impact the integration of dynamic safety stocks (the integrated model) or of the simple rule-based exogenous safety stocks (the sequential approach) has on lot-sizing. Table 4.11 presents the average, the min., the max. as well as the quantiles of 0.25, 0.5, and 0.75 of ΔTC over the randomly generated actual demand series for all combinations of the input factors. The results are based on comparison between the integrated model and the sequential approach. Additionally, under rolling horizon planning with the uncapacitated case, Table 4.12 presents the results from the comparison between the re-planning opportunity adjusted integrated model and the sequential approach. The values of α_p^* are given with respect to L and β^c. Due to the homogeneous products, α_p^* is identical for every $p \in P$. If we compare the average values to the corresponding quantile values, we can conclude the significance of the explanatory power for the average values. The min. values and in some cases the 0.25 quantiles indicate the existence of negligible noises in the comparison of the total costs with respect to an identical average realised β^c over the evaluation interval with 50 periods. To facilitate the interpretation of the results, Figure 4.5 illustrates the development of the average ΔTC if we increase L with respect to β^c in the following four comparison categories: (i) The open loop and the uncapacitated case, (ii) the open loop and the capacitated case, (iii) the rolling horizon and the uncapacitated case, and (iv) the rolling horizon and the capacitated case. Note that the corresponding average capacity utilisations for the capacitated cases and the average values of β^{RL} on which the comparison of the values of TC of the integrated model and of the sequential approach is based are presented in Table 4.9 and Table 4.10, respectively.

(i) The open loop and the uncapacitated case. We compare the integrated model with the sequential approach under the open loop in the presence of ample capacity. From Figure 4.5 (see lower right-hand corner), we notice that the integrated model outperforms the sequential approach in almost every setting. The highest average cost saving potential of the integrated model can be achieved by setting $\beta^c = 0.98$ for $L = 2$ or 3. With L=6, the impact of dynamic safety stocks declines, because a major part of the uncertainty is captured by large batch sizes.

(ii) The open loop and the capacitated case. Taking the corresponding results from Figure 4.5 (see lower left-hand corner) into account, we make the following main observations. If we set $\beta^c = 0.90$, the integrated model outperforms the

Table 4.11: The integrated model versus the sequential approach (ΔTC in %).

			uncapacitated						capacitated					
					quantile						quantile			
	β^c	L^{TBO}	avg.	min.	0.25	0.5	0.75	max.	avg.	min.	0.25	0.5	0.75	max.
open loop	0.90	1	1.2	- 4.6	- 1.3	0.9	2.5	8.2	0.5	- 1.8	- 0.2	0.1	0.7	6.9
		2	3.9	- 9.8	0.5	2.7	6.5	27.0	2.4	- 27.7	- 9.9	2.3	11.5	37.6
		3	1.2	- 14.2	- 2.4	2.4	5.3	12.3	7.0	- 9.7	1.8	7.6	11.6	22.5
		6	- 1.0	- 11.0	- 3.8	- 0.2	3.0	6.8	2.6	- 7.5	- 1.4	1.5	6.4	12.9
	0.98	1	1.2	- 11.0	- 0.6	0.7	2.1	18.9	- 1.4	- 12.6	- 1.4	- 0.7	0.1	7.1
		2	4.7	- 16.0	1.4	5.2	8.5	20.8	- 1.2	- 13.0	- 6.8	- 3.8	4.5	17.3
		3	4.4	- 11.9	1.1	3.9	7.3	19.3	9.5	- 2.2	4.7	7.2	10.8	35.1
		6	1.9	- 14.7	- 0.5	0.8	5.6	15.7	3.7	- 12.3	0.6	4.2	6.9	17.3
rolling horizon	0.90	1	0.2	- 2.0	- 0.8	- 0.0	1.3	3.5	0.8	- 1.3	0.0	0.9	1.3	4.2
		2	- 6.3	- 23.7	- 11.5	- 6.5	- 3.4	11.1	4.3	- 14.9	1.5	5.9	11.1	14.8
		3	- 3.7	- 18.3	- 8.1	- 2.9	0.8	14.3	16.1	0.7	8.5	12.6	24.4	40.7
		6	- 0.9	- 14.8	- 5.3	- 3.4	3.3	24.3	1.8	- 4.2	- 2.0	0.1	5.0	14.5
	0.98	1	0.3	- 2.8	- 1.0	- 0.6	0.9	7.9	0.6	- 1.3	- 0.2	0.6	1.3	2.8
		2	- 9.8	- 37.4	- 17.6	- 13.8	0.1	27.3	6.8	- 1.8	1.8	5.6	12.1	17.6
		3	- 15.0	- 26.8	- 24.2	- 16.9	- 11.7	7.5	9.1	- 21.6	3.3	8.2	16.9	33.3
		6	- 9.9	- 17.8	- 12.7	- 10.9	- 7.9	0.8	- 2.9	- 17.6	- 7.1	- 2.5	1.0	13.7

Table 4.12: The re-planning opportunity adjusted integrated model versus the sequential approach for the uncapacitated case under rolling horizon planning (ΔTC in %).

					quantile			
β^c	L^{TBO}	α_p	avg	min	0.25	0.5	0.75	max
0.90	1	1.00	0.2	- 2.0	- 0.8	- 0.0	1.3	3.5
	2	0.76	11.6	- 14.9	5.5	11.0	19.9	30.0
	3	0.67	6.3	- 13.9	5.3	8.5	12.7	17.8
	6	0.67	3.2	- 6.3	0.6	3.1	6.1	16.4
0.98	1	1.00	0.3	- 2.8	- 1.0	- 0.6	0.9	7.9
	2	0.67	1.9	- 19.0	- 7.6	- 1.7	4.7	54.6
	3	0.58	5.4	- 13.8	- 0.7	5.7	10.3	22.0
	6	0.51	2.8	- 4.1	0.0	0.9	5.3	16.6

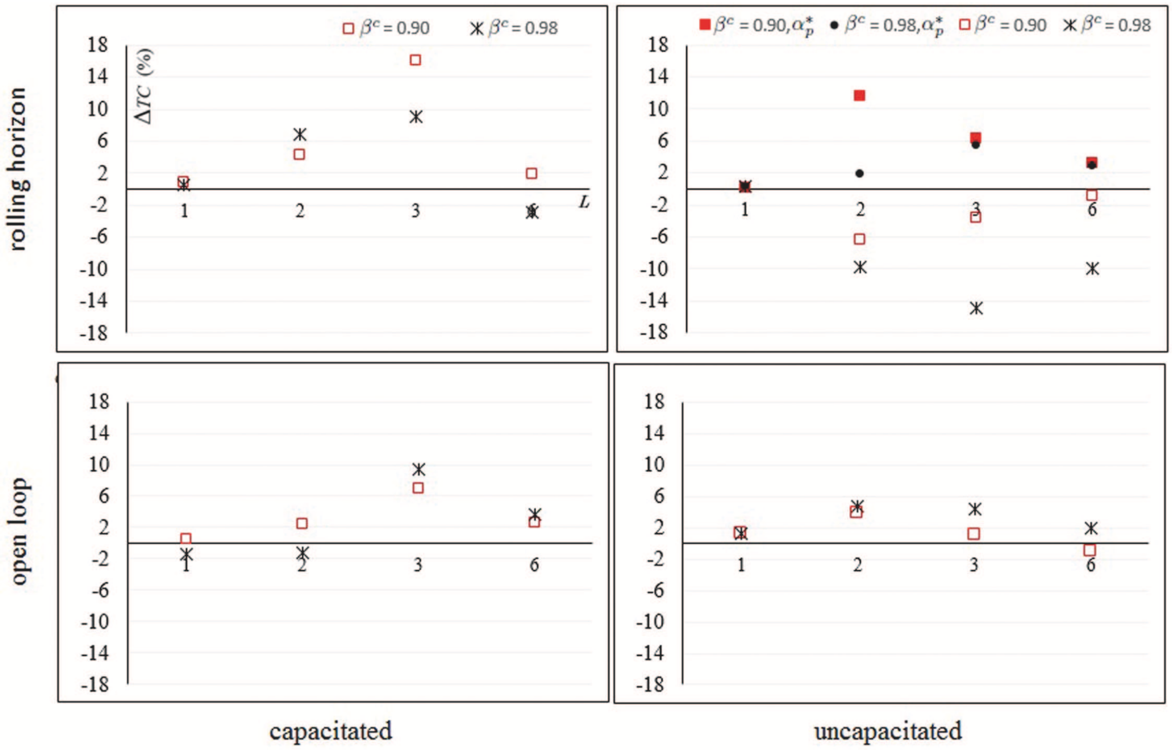


Figure 4.5: The integrated model and the re-planning opportunity adjusted integrated model versus the sequential approach: The development of the average of ΔTC in %.

sequential approach. With $\beta^c = 0.98$, the integrated model has, on average, cost-saving advantages against the sequential approach if L = 3 or 6. However, if L = 1 or 2, then the sequential approach can outperform the integrated model. These observations can be explained as follows. Under a capacitated case, the production decisions may not merely follow the stochastic demand process. After all, the inventories can be built during the earlier periods and then be used during the subsequent periods of peak demand. This can lead to a higher uncertainty when we anticipate the actual inventory levels at the beginning of the subsequent replenishment cycles (the expected initial inventories). A longer frozen horizon (here $T^{\mathrm{fh}} = 10$) can include a higher number of replenishment cycles (particularly if $L = 1$). Thus, the anticipation of the actual initial inventory levels of the replenishment cycles involves more uncertainty towards the end of the frozen horizon. As our results show, the sequential approach will not always outperform the integrated model. However, for some settings, the sequential approach with simple-rule exogenous safety stocks can provide better results than the integrated model.

(iii) The rolling horizon and the uncapacitated case. If we take the integrated model into account, we notice that, with the exception of $L = 1$, the sequential approach can outperform the integrated model with considerable average TC differences. These (unexpected) observations can be explained as follows. In the presence of ample capacity, the rolling horizon provides a high degree of flexibility for re-planning when it comes to reacting to demand uncertainty. In general, such a high re-planning opportunity allows for the fulfilment of possible backorders with a maximum delay of one period. Furthermore, as L increases, more demand is bundled into one lot. Such a large lot size can easily cover the demand uncertainty observed during one period under the rolling horizon. Thus, it is no longer necessary to add extra safety stocks. However, in the case of the integrated model, if $L > 1$, the dynamic safety stock is calculated by taking the demand uncertainty of the entire replenishment cycle into account, which leads to an excess inventory. If we incorporate re-planning opportunity adjustment, we notice that the integrated model can outperform the sequential approach in every input factor combination. The re-planning-opportunity adjusted integrated model can prevent excess dynamic safety stocks and lead to a considerable improvement in performance.

(iv) The rolling horizon and the capacitated case. In our last category, we compare the performance of the integrated model with that of the sequential approach under the rolling horizon and the capacitated case. In Figure 4.5, we notice that the integrated model can outperform the sequential approach in almost every case. The only exception is $\beta^c = 0.98$ with $L = 6$. Note that this setting is not really capacitated, as Table 4.9 shows an average capacity utilisation of 47.8%. This is an interesting observation and not in accordance with the results we obtained under the rolling horizon in the presence of ample capacity without the re-planning opportunity adjustment ($\alpha_p = 1$). It can be explained as follows. The restricted flexibility for re-planning caused by capacity limitation implies that we cannot react to demand uncertainty in less time than one period. Thus, it is necessary to take the demand uncertainty over entire lengths of the replenishment cycles into account, especially those that are longer than the length of the frozen horizon at the beginning of the planning horizon. Due to the capacity restriction, there is less likelihood that the replenishment cycles will be affected (changed) by future planning revisions. Thus, the dynamic safety stocks placed by the integrated model can lead to a better performance than the sequential approach, which ignores the length of the replenishment cycles.

We investigated other cases that we do not report in detail because of limited space. Our numerical study shows that, if we freeze the schedule of the entire planning horizon ($T^{\text{fh}} = 12$), the results do not vary much from the setting with $T^{\text{fh}} = 10$. In our numerical study, we presented the results with the assumption that $c_p^{\text{ds}} = 0.2$. We also looked into other penalty cost coefficients for the deviation from the target (dynamic) safety stocks, i.e., $c_p^{\text{ds}} = \{0.1, 0.3, 1\}$. Similar results are obtained for $c_p^{\text{ds}} = \{0.1, 0.3\}$. However, using identical deviation cost coefficients for both components of the order-up-to-level ($c_p^{\text{ds}} = 1$) usually leads to worse results than those we get under different cost coefficient values.

Integrated model versus stochastic dynamic program

We evaluate the absolute performance of the integrated model under rolling horizon planning by means of the theoretical lower bound obtained from the SDP. We take both the capacitated and uncapacitated cases, as well as $\beta^c = \{0.90, 0.98\}$ with $L = 1$, into account. The comparison is based on 200 independent replications of the realised demand series, which is a sufficient number to obtain statistically significant comparison results.

We use the following comparison procedure. First, we solve the integrated model and obtain the average realised service level (β^{RL}) for all independent replications of the realised demand series. Then, we solve the SDP with different backlog penalty costs (c_p^{bl})

and select a value that results in the SDP returning the same β^{RL} as the one obtained from the integrated model. Afterwards, we calculate the total cost difference ($\Delta TC = TC^{\text{IM}} - TC^{\text{SDP}}$) for each independent replication of the actual demand series. We also conduct a 99% t-paired confidence interval test to prove the statistical significance of the average ΔTC.

Table 4.13 presents the average and standard deviations of β^{RL} and TC over all independent replications of the actual demand series for both the integrated model and the SDP. Table 4.13 further shows the target cycle fill-rate for the integrated model and the selected c_p^{bl} value for the SDP.

Taking the uncapacitated case into account, we see that the average ΔTC increases as we move from $\beta^c = 0.90$ to $\beta^c = 0.98$. The integrated model with $\beta^c = 0.98$ results in excess inventories under the rolling horizon.

If we take the capacitated case into account, we note that the integrated model returns higher average TCs than the SDP for both $\beta^c = 0.90$ and $\beta^c = 0.98$. Due to its full look-ahead capability, the SDP can use the available capacity to better prepare for the upcoming low or high demand periods than the integrated model.

Table 4.13: Integrated model versus stochastic dynamic program under rolling horizon planning.

	Integrated Model			SDP	
		β^{RL}	TC	β^{RL}	TC
	β^c	avg.[St.Dev.]	avg.[St.Dev.]	avg.[St.Dev.]	avg.[St.Dev.]
Uncap.	0.90	95.0 [0.8]	268.0 [20.8]	95.0 [0.8]	267.8 [20.8]
	0.98	99.1 [0.4]	705.4 [15.4]	99.1 [0.5]	700.1 [15.2]
Cap.	0.90	95.1 [0.7]	268.9 [20.3]	95.1 [0.7]	264.4 [19.7]
	0.98	99.0 [0.5]	726.9 [24.0]	99.0 [0.5]	711.7 [21.2]

4.6 Conclusions

We addressed a stochastic capacitated lot-sizing problem by introducing a stochastic dynamic program and three MILPs, i.e., the sequential approach, the new integrated model, and the adjusted integrated model to deploy capacity flexibilities under rolling horizon planning.

In the experimental study, we first compared the performance of the integrated model to the widely-used sequential approach. Figure 4.6 summarises our main findings from the comparison of these modelling approaches. According to our experimental study, we found that, if capacity is binding, the integrated model leads to lower total costs than the sequential approach for an identical average realised service level under rolling horizon planning. However, we noticed that, if capacity is not binding, the integrated model leads to excess dynamic safety stocks since it ignores re-planning opportunities to exploit capacity flexibilities under rolling horizon planning. The new adjusted integrated model could outperform the sequential approach since it was able to reduce dynamic safety stocks by correctly anticipating re-planning opportunities. The comparison under open loop planning showed that, if capacity is not binding, the integrated model returned lower total costs than the sequential approach for an identical realised service level but it was not always true if capacity is limited.

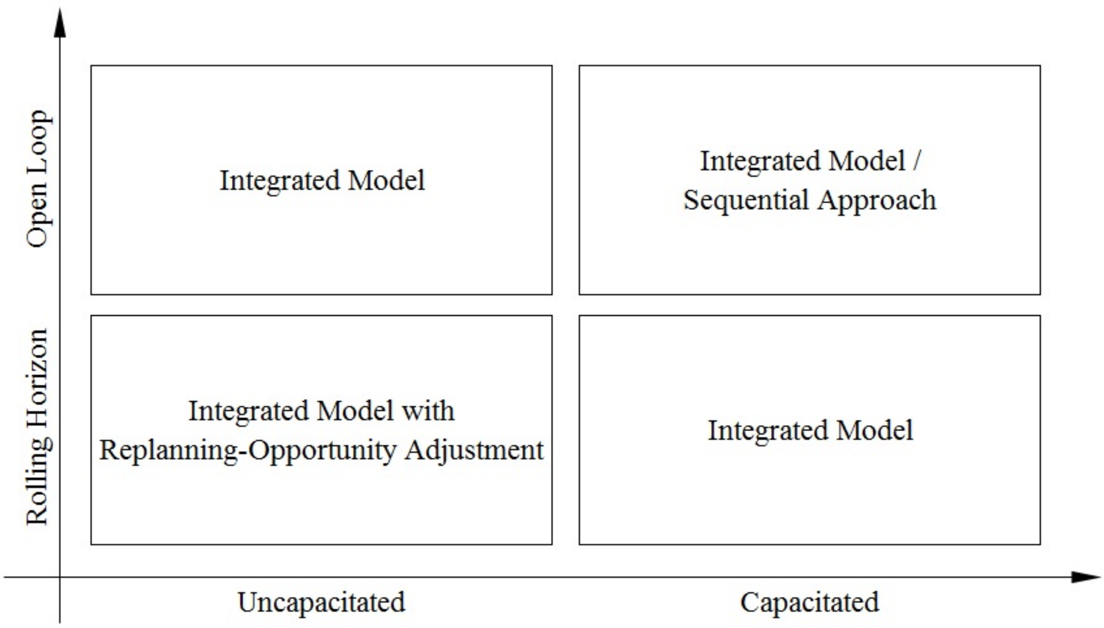


Figure 4.6: Choosing an appropriate modelling approach with respect to the planning approach and capacity level.

The comparison of the integrated model with the SDP showed that, under the uncapacitated case, the integrated model can provide near optimal solutions if the target service level is low. However, the absolute performance of the integrated model decreases if we set a high target service level because the integrated model yields excess safety stocks. If capacity becomes binding, the absolute performance of the integrated model reduces due to the lack of a full look-ahead capability that would have allowed the anticipation of low and peak demand periods for an optimal assignment of the limited capacity.

Chapter 5

Integrated Lot-sizing, Scheduling and Dynamic Safety Stock Planning: A Real-World Case Study

We develop and apply a new approach for integrated production lot-sizing and safety stock planning under serially-correlated demands. In the literature, as well as in most Advanced Planning Systems (APS), stochastic lot-sizing and scheduling optimisation problems are often translated into a deterministic problem by entering predetermined safety stock targets into the production planning model. The sequential approach represents an approximation because the safety stock requirement depends on the production plan, which is not known a priori. Ideally, both problems should be addressed simultaneously in an integrated model. We develop a Mixed-Integer Linear Program (MILP) that endogenously determines dynamic safety stocks over replenishment cycles with continuous and non-equidistant lengths according to the uncertainty parameters for both uncorrelated and serially-correlated demands. To account for re-planning opportunities under rolling horizon planning, we introduce a new method that prevents an excessive build-up of safety stocks. We use the integrated model to help our industrial partner from the process industry with the quantification of the cost-saving potential over the traditional sequential approaches. Based on a real-world dataset, we find that feeding cost-optimised safety stock targets into the sequential approach instead of employing a widely-used Rule-of-Thumb (RoT) can already deliver cost savings of up to 10%. Further cost improvements of up to 20% over the cost-optimised safety stock targets can be obtained by the integrated model. The integrated approach can provide an additional 5% in cost saving if it takes serially-correlated demand into account. Despite its complexity, the integrated model is also appealing for practical purposes because it produces robust highly promising results and does not require any separate safety stock specification upfront.

5.1 Introduction

The acquisition of effective production plans that provide high customer service at a low cost plays a critical role in many industries including, e.g., the consumer products, electronics, paper, pharmaceutical, steel and the chemical industry (Jans and Degraeve, 2008). Many use planning software such as MRP II and APS. Despite the application of these tools, the various production constraints in real-world settings and the uncertain nature of the product demands still render this a challenging task to achieve effective production plans. In the concrete application example from the chemical company that we study in this chapter, we face a limited production capacity, minimum production quantities and times, sequence-dependent setup costs and times as well as uncertain and serially-correlated demands.

Our industrial partner is one of the world's largest chemical companies and has more than 350 production plants worldwide. In order to facilitate and improve its production planning activities within and across plants, the company has started with a widespread implementation of APS for production planning in its plants several years ago. Although many plants have previously created their production plans manually with the help of self-developed spreadsheets, most of them have APS in place nowadays. The extent to which the plants use the system's available functionality for production planning varies considerably, however. Some plants only make use of the graphical planning board (instead of the previously used spreadsheet), but still create the actual production plan manually. Others exploit the planning heuristics that the system offers, for instance part period balancing or the Groff heuristic (Groff, 1979). Other plants work with APS' mathematical programming model to optimise their production plans.

This broad range of sophistication in terms of planning, referred to as planning maturity levels by our industrial partner, suggests a clear potential for further improvements. While the development path and potential for the plants at the low maturity levels is rather obvious, it is not apparent to the company management whether they should also invest in trying to push for further improvements at the highly mature plants that already use an optimisation model.

The reason why the management team considers the latter option at all is the awareness that even the optimisation model now available in APS has certain shortcomings. As is common for most APS, the simultaneous lot-sizing and scheduling problem with uncertain demands is broken down into two planning problems that are solved sequentially: (1) A safety stock planning problem that addresses the demand uncertainty and (2) a deterministic production planning problem that uses the predetermined safety stocks as inputs. The production and safety stock planning problems are interrelated, however. The safety stock needs to buffer against the demand uncertainty over the replenishment cycle. In the stochastic-demand inventory literature, the replenishment cycle comprises of the replenishment lead time plus the review interval. In the production planning context, it is the time span between two consecutive production lots of a product.

Figure 5.1 illustrates a production plan for one product over twelve time periods. Production occurs in periods 1 and 6 and is marked red. In the remaining periods (marked blue), this product is not produced. Therefore, the first replenishment cycle extends over 5 periods, from the beginning of period 1 to the end of period 5. In order to decide about the production quantity in period 1, we need to take into account the demands of periods 1 to 5 (including), because new material will only become available in period 6 when next the lot is produced. If the next lot has already been produced in period 4 instead of in period 6, the replenishment cycle is 3 periods. Consequently, the production plan clearly impacts the safety stock requirement. On the other hand, the safety stock quantities need to be produced at some point. Therefore, the safety stock requirement also affects the production planning problem.

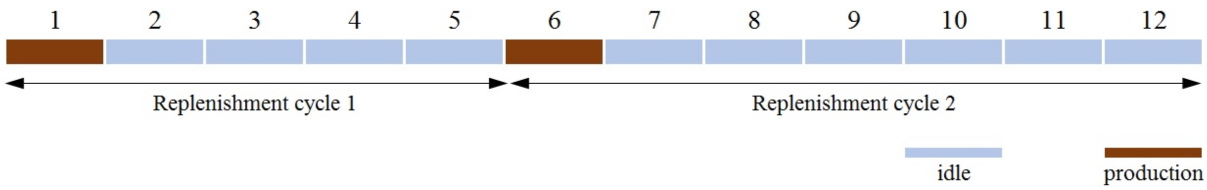


Figure 5.1: Replenishment cycle.

The open questions asked by our industrial partner for the plants at the high maturity level refer to:

1. The appropriateness of the predetermined safety stocks that the plants currently use in the sequential safety stock and production planning approach.

2. The potential savings achieved by using a simultaneous approach in the form of an

integrated (stochastic) production and safety stock planning model.

While some integrated (stochastic) production and safety stock planning models exist in the literature, none of them addresses the Stochastic General Lot-sizing and Scheduling Problem (S-GLSP) that describes a real-world application setting like ours (see literature review in Chapter 2 for details). Moreover, the company under study and many other companies in practice pursue a rolling-horizon planning approach. The existing models, however, ignore the rescheduling opportunity available under rolling-horizon planning when the required safety stocks are determined. If there is excess production capacity, it might be beneficial to reduce a product's safety stock and save inventory holding costs by taking advantage of the possibility to pre-pone the production of the next lot where necessary. In order to reflect this rescheduling opportunity and its anticipation accurately, a stochastic dynamic programming model needs to be used (e.g., see Section 4.3 in the previous chapter). Such an approach suffers from the curse of dimensionality, however. Consequently, it is of only limited use in practice. For practical purposes, it is desirable to find a way to consider the rescheduling flexibility in a mixed-integer programming model formulation.

In our search for answers to the questions raised by our industrial partner, we actually make the following contributions:

- From a modelling and methodology point of view, we introduce an MILP for the stochastic lot-sizing and scheduling problem that determines the exact length of a product's replenishment cycle and thus enables the endogenous sizing of the required safety stocks based on the demand uncertainty parameters during the replenishment cycle. The demand uncertainty parameters during the replenishment cycles are determined in the presence of serially-correlated demands. We develop a bivariate linearisation technique to approximate non-linear dynamic safety stock requirements. We present a new method to adjust the safety stock levels, thus taking advantage of the flexibility provided through the regular re-scheduling opportunities under rolling horizon planning. We propose a period-based decomposition approach for obtaining a promising solution in a reasonable amount of time since even the deterministic lot-sizing and scheduling problem is known as an NP-hard problem.

- We explore the cost-saving potential that results from the use of more sophisticated modelling approaches based on a real-world dataset from the chemical company under study from a managerial point of view. We find that, even if a high level of sophistication in terms of production planning has already been achieved in the form of the common sequential approach, the optimisation of the exogenous safety stock targets can provide significant gains. Compared to a widely-used RoT at the company, the implementation of cost-optimised Days-of-Supply (DoS) safety stock targets according to a stochastic-demand inventory control logic shows a cost-saving potential of up to 10% for both a common Days-of-Supply (cDoS) target for all products and individual Days-of-Supply (iDoS) targets in the studied scenarios. In all settings, further improvements of up to 20% over the cost-optimised DoS safety stock targets can be obtained with our newly developed integrated production and safety stock planning model. If we take serially-correlated demand into account, further improvement up 5% can be achieved. Even though the integrated models are more complex, it is appealing, also from a practical point of view, to apply them because they are robust and the planner is not required to specify the replenishment cycle length for the safety stock determination, since this is all done endogenously.

The remainder of this chapter is organised as follows. Section 5.2 outlines the real-world planning problem of the company under study and its current planning practice. Section 5.3 presents the general problem description, the assumptions and notation. Section 5.4 shows how we estimate demand uncertainty. Section 5.5 presents the mathematical model formulations for the sequential approaches. In Section 5.6, we develop the integrated approach. Section 5.7 provides the solution approaches. Section 5.8 compares the performance of the different planning approaches numerically and Section 5.9 concludes this chapter.

5.2 The Production Setting and Current Planning Practice of the Company under Study

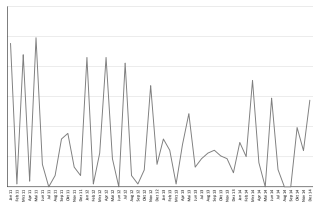
The production plant under consideration is a medium pressure multi-purpose system. The input materials for the chemical reaction are provided through pipelines in any quantity required. Therefore, we can assume ample supply. Through different reaction variants, seven final products can be produced. Six of them are sold to external customers, one is targeted solely for internal usage at another plant. Due to the special organisational structure of the company, it is common to have not only external, but also internal customers. Whenever a product is to be produced, a setup time of 3 days, as well as a sequence-dependent setup cost of several tens of thousands of Euros is incurred. Due to technical and quality restrictions, a minimum lot size of 14 days of production is required for each product. The output products are stored in tanks with different volumes holding between 140 and 510 tons. If the product demand cannot be fully satisfied from stock, excess demand is backlogged.

Demands for the different products fluctuate over time (see Figure 5.2). According to the company's current planning guidelines, the creation of a forecast is generally required for all products with external demand. Therefore, for six of the seven products the marketing and sales department creates monthly forecasts and forwards the numbers to the production planner via APS. Each month the planner receives updated forecasts for the next 12 months. For the internal Product 7, no forecast is generated. For this particular product, the production planner can align the production plan very well with the downstream plant (internal customer). Information is exchanged on a regular basis. Therefore, it can be assumed that there is no uncertainty at all attached to this product demand, neither with respect to timing nor to quantity.

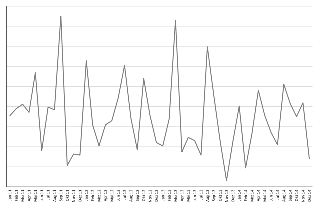
The forecast data form the basis for the generation of a new production plan each month. The planner feeds these data into the mathematical optimisation module of APS. Before starting the optimisation run, the inputs can be manually modified, if required. For instance, if a bias in the forecasting process has been observed in the past, the planner can try to correct it. Ideally, the correction should take place at the source where the forecast is generated. However, the power to do so lies with the marketing and sales department, and thus outside the planner's discretion. The planner can only "correct" it at a later stage, which he controls, namely right before the production plan is created. This situation is similar to the one described in Manary and Willems (2008).

The planner uses the standard APS optimisation model, which is a deterministic lot-sizing and scheduling model formulation with a predetermined safety stock target for each product, to generate the production plan for the upcoming 12 months. The plan for the next month gets implemented and a new plan is created in the upcoming month based on the realised demand and the updated forecast data. This represents a rolling horizon planning approach with a planning horizon of 12 months and a re-planning interval of 1 month.

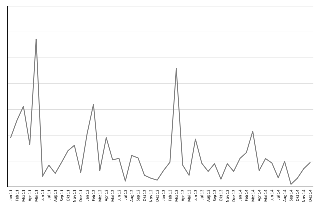
The challenge for the planner lies in the specification of appropriate safety stock targets for the different products, which are required as inputs to the optimisation model. Currently, he uses a simple RoT. He works with a DoS approach, as recommended by



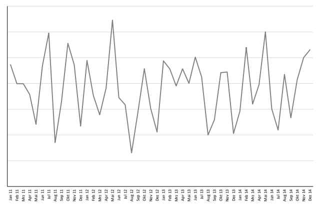
(a) Product 1



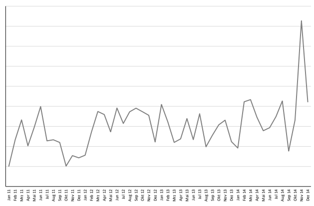
(b) Product 2



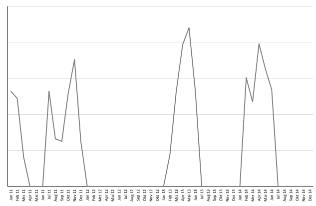
(c) Product 3



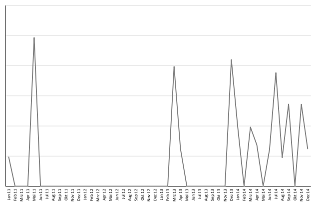
(d) Product 4



(e) Product 5



(f) Product 6



(g) Product 7

Figure 5.2: Historical demand time series for the different products.

many software vendors (see Neale and Willems, 2015). Since the production plan can be adjusted every month, the planner chooses a common DoS target of one month, i.e., 30 days, as the safety stock target for each product.

Given this planning setting, the goal of this chapter is two-fold. First, we seek to assess the appropriateness of the safety stock specification with the current RoT. To this end, we compare it with an approach that sets the safety stock targets based on cost optimisation considerations as is common in stochastic-demand inventory theory. Second, we evaluate the cost-saving potential resulting from a transition to an integrated (stochastic) production and safety stock planning model that endogenously determines the safety stock requirements.

Since we do not have access to the optimisation model in the company's APS, we rebuild it as closely as possible. For our analysis we parametrise it with the two different exogenous safety stock specifications (RoT vs. cost-optimal). The advantage of using our own deterministic model formulation in the sequential approach is that we can clearly identify the value added by the safety stock integration in a next step, because we can ensure that both models only differ in terms of the safety stock aspect. Such an assessment would not be straightforward in case of the APS optimisation model. In the following sections we outline our mathematical model formulations.

5.3 General Problem Description, Assumptions and Notation

We consider a single machine that can produce multiple products $p \in \mathcal{P} = \{1, ..., |\mathcal{P}|\}$. We formulate the lot-sizing and scheduling problem as a tailored hybrid GLSP model that combines large-bucket and small-bucket time scales. The finite planning horizon is divided into equidistant macro periods $t \in \mathcal{T} = \{1, ..., |\mathcal{T}|\}$ (large buckets), representing months and continuous, non-equidistant micro periods $s \in \mathcal{S} = \{1, ..., |\mathcal{S}|\}$ (small buckets). Multiple products can be produced within a macro period, but only a single product is produced in every micro period. $\mathcal{S}_t \subset \mathcal{S}$ denotes the set of micro periods within the macro period t, and set $\mathcal{S}^{\text{end}} \subset \mathcal{S}$ summarises the last micro periods of all the macro periods. Moreover, $s_t^{\text{last}} \in \mathcal{S}^{\text{end}}$ represents the last micro period of macro period t, and t_s returns the corresponding macro period of micro period s. The starting time of the planning horizon is given as $\boldsymbol{\tau}_0(= 0)$. The production capacity consumption (in time per unit) of product p is given by coefficient k_p^{prd}. The available production capacity is defined per macro period in terms of time, K_t^{prd}. There is a minimum production quantity q_p^{min} per product and a minimum production time z^{minq}, which is identical for all products. All processed products are stored in dedicated tanks with a product-specific limited storage capacity K_p^{inv}. The storage capacity consumption per unit of product p is given as k_p^{inv}.

The daily demand of each product p is random, non-stationary variable and follows a normal distribution with the mean μ_p and the standard deviation σ_p. The probability density and cumulative distribution functions of the standard normal distribution are indicated by $\phi(z_p)$ and $\Phi(z_p)$ respectively. We assume the demand of each product to be independent of the other products, but may be serially-correlated over time. Each processed item is available to satisfy demand as soon as it leaves the production process however we do not have to wait for the entire production lot to be completed. Unsatisfied demand is backlogged.

The goal is to size the safety stock targets and derive a solution in terms of lot-sizes and detailed schedules that minimises the total costs. The total costs consist of four components: (1) A linear inventory holding cost per product, c_p^{h}, (2) a backlogging cost, c_p^{bl}, which corresponds to the product's contribution margin, (3) a sequence-dependent setup cost for switching from product p to m, c_{pm}^{setup}, (as well as a setup time z_{pm}^{setup}) and (4) a penalty cost for deviations from the safety stock target of product p, c_p^{sst}. The latter cost component is only used in the sequential approach. In integrated approaches, we penalise the expected backlogged quantities at the end of the replenishment cycles.

Since we consider using a rolling-horizon planning approach, we might have to fix the production quantity of a specific product right at the start of the new planning horizon in order to fulfil the minimum production quantity constraint from the previous planning run, given as q_p^{initial}. Moreover, for each planning run, we know the initial inventory quantity of each product, y_{p0}, and the initial setup configuration on the machine, δ_{p0}.

5.4 Calculation of Forecast Bias and Demand Uncertainty

In this section, we try to answer the question how to estimate the forecast error for each product. This is what we need to buffer against with safety stock. We use an out-of-sample evaluation approach. We divide the provided demand and forecast time series into an estimation phase and an evaluation phase. In the estimation phase, $t \in \mathcal{E} = \{1, \ldots, |\mathcal{E}|\}$, we compute the forecast uncertainty and the forecast bias. In the evaluation phase, $t \in \mathcal{V} = \{|\mathcal{E}| + 1, \ldots, |\mathcal{E}| + |\mathcal{V}|\}$, we conduct the numerical comparison between the different production planning approaches, which we develop in this chapter. We pretend that this time span (for which we already know the actual demand realisations) represents the future, so that we can simulate the performance of the different models under real-world conditions.

For the computation of the forecast measures in the estimation phase, we follow Cachon and Terwiesch (2012), Chapter 12. In the interest of an easy presentation, we drop the product index p in the following exposition. For each forecast step $i \in \mathcal{F} = \{1, \ldots, |\mathcal{F}|\}$, i.e., for all 1-month ahead forecasts, 2-month ahead forecasts and so on, we compute the mean and standard deviation over all months in the estimation phase, $t \in \mathcal{E} = \{1, \ldots, |\mathcal{E}|\}$, of the ratio between the actual demand in period t (indicated by a_t) and the forecast value (indicated by $f_{t-i,t}$) for this month as it was generated i months in advance:

$$Mean(DF_i) = \frac{1}{|\mathcal{E}|}\sum_{t=1}^{|\mathcal{E}|}\frac{a_t}{f_{t-i,t}}, \qquad \forall i \in \mathcal{F} \tag{5.1}$$

$$SD(DF_i) = \sqrt{\frac{1}{|\mathcal{E}|-1}\sum_{t=1}^{|\mathcal{E}|}\left(\frac{a_t}{f_{t-i,t}} - Mean(DF_i)\right)^2}. \qquad \forall i \in \mathcal{F} \tag{5.2}$$

In case of a perfect forecast for step i, $Mean(DF_i) = 1$. If $Mean(DF_i)$ is larger (smaller) than 1, there is a negative (positive) forecast bias. By multiplying the respective mean and standard deviation with the future forecast for a particular month in the evaluation phase and beyond, i.e., periods $t = |\mathcal{E}|+1, \ldots, |\mathcal{E}|+|\mathcal{V}|+|\mathcal{F}|-1$, we convert the relative forecast accuracy measure into an absolute one. (Note that the time index t runs until time period $|\mathcal{E}|+|\mathcal{V}|+|\mathcal{F}|-1$, because at the end of period $|\mathcal{E}|+|\mathcal{V}|$, we still need to create a production plan based on the forecasts for the next $|\mathcal{F}|-1$ months so that the production in $|\mathcal{E}|+|\mathcal{V}|$ is planned based on the 1-month ahead forecasts.) Thus, we obtain the bias corrected point forecast (forecast mean) that we use for the MILP models in the sequential and integrated approaches as

$$\tilde{f}_{t-i,t} = Mean(DF_i) \cdot f_{t-i,t}, \qquad \forall i \in \mathcal{F}, \quad t = |\mathcal{E}|+1, \ldots, |\mathcal{E}|+|\mathcal{V}|+|\mathcal{F}|-1 \tag{5.3}$$

and the standard deviation that represents the forecast uncertainty used in the integrated model and for setting the safety stock targets in the first step of the sequential model as

$$\tilde{\sigma}^f_{t-i,t} = SD(DF_i) \cdot f_{t-i,t}, \qquad \forall i \in \mathcal{F}. \quad t = |\mathcal{E}|+1, \ldots, |\mathcal{E}|+|\mathcal{V}|+|\mathcal{F}|-1 \tag{5.4}$$

Based on the $Mean(DF_i)$ ratios, we find that the forecasts are basically not biased at all for three out of all the products 1-5. Of the other two products, one exhibits a rather strong positive forecast bias and the other a strong negative forecast bias. For product 6, we cannot perform this kind of analysis on an individual forecast-step basis because of the special demand structure with many zero-demand periods. The remaining number of periods with positive demand realisations is too small for a sound statistical analysis of every individual forecast step. Therefore, we only compute the $Mean(DF)$-ratio across all positive historical demands and forecasts. We detect no bias for this product, either.

5.5 Lot-sizing and Scheduling with Exogenous Safety Stock Targets

Under the sequential modelling approach, we first determine safety stock targets for all products. Next, we use them in a deterministic lot-sizing and scheduling model.

5.5.1 Rule-of-Thumb for Safety Stock Targets

For each product p we determine a safety stock target for macro period (month) t, y_{pt}^{sst}. The planner's current RoT uses a cDoS target of one month (30 days) for safety stock sizing. Since the re-planning interval is one month, having a safety stock of one month was considered to be reasonable by the planner. Accordingly, when creating the production plan for the upcoming 12 periods, $t = 1, \ldots, 12$, at the end of period 0, the planner considers the available demand forecasts for these periods at this point in time denoted as $\tilde{f}_{p,0,t}$ for each product p (see Section 5.4 for details on $\tilde{f}_{p,0,t}$). Since one period corresponds to one month in our model, the safety stock target of product p in period t is set equal to the forecast demand of that period in order to satisfy the DoS target, i.e., $y_{pt}^{\text{sst}} = \tilde{f}_{p,0,t}$, for every $t \in \mathcal{T}$.

5.5.2 Cost-based Approach for Safety Stock Targets

The RoT does not take any costs into account. In stochastic-demand inventory models, a cost-based approach is usually suggested for the safety stock determination. As an alternative, we use the (s, Q) optimisation approach by Silver et al. (2017) that determines the reorder point s (hereafter s^{ROP}) under a given lot-size Q (hereafter $q_p^{\min}$) based on an approximate total cost minimisation under stationary normally distributed demand, which is commonly known and applied in practice. In order to account for non-stationary demand, APS vendors recommend the use of a DoS measure, instead of a constant safety stock target quantity (see, e.g., Neale and Willems, 2015). Even though the DoS target is kept constant, such an approach self-adjusts the safety stock target quantity with the (non-stationary) demand without adding complexity. We convert our determined safety stock target quantity into a DoS measure and consider this in the specification of y_{pt}^{sst}.

Let $\bar{c}_p^{\text{setup}}$ denote the average fixed setup cost of product p.[1] L_p denotes the lead time,

[1]Even though we face sequence-dependent setup costs in our production setting, we do not need to worry about this in the specification of the cost model because $\bar{c}_p^{\text{setup}}$ is not relevant for determining the

which corresponds to the replenishment cycle in our production planning context. In our production setting, we choose the lot size for each product p, q_p^{min}, equal to the daily product-specific production rate times the minimum production time z_p^{minq}, as this represents a lower bound. The expected total cost $C_p(z_p)$ of a specific product p per period is (see Silver et al., 2017)

$$C_p(z_p) = \frac{\mu_p}{q_p^{\mathrm{min}}} \cdot \bar{c}_p^{\mathrm{setup}} + c_p^{\mathrm{h}} \cdot \left(\frac{q_p^{\mathrm{min}}}{2} + z_p \cdot \sigma_p \cdot \sqrt{L_p} \right) + \frac{\mu_p}{q_p^{\mathrm{min}}} \cdot c_p^{\mathrm{bl}} \cdot \sigma_p \cdot \sqrt{L_p} \cdot G_p(z_p), \quad (5.5)$$

with $G_p(z_p) = \phi(z_p) + z_p \cdot (1 - \Phi(z_p))$ and $z_p = \frac{x - \mu_p}{\sigma_p}$ for any $x \in \mathbb{R}$. The resulting optimality condition is $\Phi(z_p) = 1 - c_p^{\mathrm{h}} \cdot q_p^{\mathrm{min}} / c_p^{\mathrm{bl}} \cdot \mu_p$. Given the optimal z_p, the safety stock for product p, $\bar{y}_p^{\mathrm{sst}}$, follows as $\bar{y}_p^{\mathrm{sst}} = z_p \cdot \sigma_p \cdot \sqrt{L_p}$ and the reorder point as $s_p^{\mathrm{ROP}} = L_p \cdot \mu_p + \bar{y}_p^{\mathrm{sst}}$.

With respect to the demand parameters, we update μ_p and σ_p at the beginning of each year only, based on the available forecasts and their standard deviations at corresponding point in time. Since demand is serially correlated, σ_p is determined by taking covariances between two consecutive forecast values into account. Recall from Section 5.4 that $\tilde{f}_{p,i,t}$ and $\tilde{\sigma}_{p,i,t}^{f}$ indicate the point forecast and its standard deviation for product p, which is created in period i for period t, respectively. Let $cov(\tilde{f}_{p,0,t}, \tilde{f}_{p,0,t+1})$ denote the covariance between two consecutive point forecasts $\tilde{f}_{p,0,t}$ and $\tilde{f}_{p,0,t+1}$. Assuming that $t = 1$ indicates January and one month = 30 days we calculate:

$$\mu_p = \frac{1}{12 \cdot 30} \sum_{t=1}^{12} \tilde{f}_{p,0,t}, \quad (5.6)$$

$$\sigma_p = \frac{1}{12 \cdot 30} \sum_{t=1}^{11} \sqrt{(\tilde{\sigma}_{p,0,t}^{f})^2 + (\tilde{\sigma}_{p,0,12}^{f})^2 + 2 \cdot cov(\tilde{f}_{p,0,t}, \tilde{f}_{p,0,t+1})}. \quad (5.7)$$

The only parameter left to specify is L_p, the replenishment cycle. This represents the actual challenge because it depends on the production plan that we do not know when determining the safety stock targets. We can only use an approximation. Since we can re-plan every month, we choose $L_p = 30$ days for all products.

Given this parameter specification, we can determine an optimal z_p value and consequently safety stock target quantity $\bar{y}_p^{\mathrm{sst}}$ for each product. By dividing $\bar{y}_p^{\mathrm{sst}}$ by μ_p we obtain

optimal safety stock level.

iDoS targets ($\bar{y}_p^{\text{iDoS}}$):

$$\bar{y}_p^{\text{iDoS}} = \frac{\bar{y}_p^{\text{sst}}}{\mu_p}. \qquad \forall p \in \mathcal{P} \tag{5.8}$$

If a cDoS target for the different products is desired (indicated by $\bar{y}^{\text{cDoS}}$), it follows that

$$z_p \cdot \sigma_p \cdot \sqrt{L_p} = \mu_p \cdot \bar{y}^{\text{cDoS}} \Leftrightarrow z_p = \frac{\mu_p}{\sigma_p \cdot \sqrt{L_p}} \cdot \bar{y}^{\text{cDoS}}. \qquad \forall p \in \mathcal{P} \tag{5.9}$$

By inserting (5.9) into (5.5) and summing up over all products we obtain the relevant total cost function $C(\bar{y}^{\text{cDoS}})$ and the first order condition for the optimal $\bar{y}^{\text{cDoS}}$ value:

$$C(\bar{y}^{\text{cDoS}}) = \sum_{p=1}^{|\mathcal{P}|} C_p(z_p), \tag{5.10}$$

$$\frac{dC(\bar{y}^{\text{cDoS}})}{d\,\bar{y}^{\text{cDoS}}} = \sum_{p=1}^{|\mathcal{P}|} \left(c_p^{\text{h}} + \frac{\mu_p \cdot c_p^{\text{bl}}}{q_p^{\min}} \cdot \left(\Phi\left(\frac{\mu_p}{\sigma_p \cdot \sqrt{L_p}} \cdot \bar{y}^{\text{cDoS}} \right) - 1 \right) \right) \frac{\mu_p}{\sigma_p \cdot \sqrt{L_p}} = 0. \tag{5.11}$$

We can simply use a numerical root search method to find the the optimal value of $\bar{y}^{\text{cDoS}}$.

For $\bar{y}^{\text{DoS}} \in \{\bar{y}_p^{\text{iDoS}}\ ,\ \bar{y}^{\text{cDoS}}\}$ we obtain y_{pt}^{sst} as (where $\sum_a^b x = 0$ if $a > b$):

$$n^{\text{month}} = \left\lfloor \frac{\bar{y}^{\text{DoS}}}{30} \right\rfloor \quad , \qquad n^{\text{day}} = \bar{y}^{\text{DoS}} \mod 30, \tag{5.12}$$

$$y_{pt}^{\text{sst}} = \underbrace{\sum_{i=t}^{t+(n^{\text{month}}-1)} \tilde{f}_{p,0,i}}_{\text{full months}} + \underbrace{n^{\text{day}} \cdot \frac{\tilde{f}_{p,0,t+n^{\text{month}}}}{30}}_{\text{remaining days}}. \tag{5.13}$$

5.5.3 Mixed-Integer Linear Program for Lot-sizing and Scheduling with Exogenous Safety Stocks

We enter the predetermined safety stock targets, y_{pt}^{sst}, into the following deterministic lot-sizing and scheduling model to determine the production plan for the planning horizon $t = 1, \ldots, |\mathcal{T}|$. Let d_{pt} correspond to the (bias corrected) point forecasts $\tilde{f}_{p,0,t}$ that have been created at the beginning of the planning horizon for periods $t = 1, \ldots, |\mathcal{T}|$ (see

Section 5.4). The notation is summarised in Table 5.1.

$$\min TC =$$

$$\sum_{p\in\mathcal{P}}\sum_{t\in\mathcal{T}} c_p^{\mathrm{h}} \cdot \mathbf{y}^{+}_{p,s_t^{\mathrm{last}}} + \sum_{p\in\mathcal{P}}\sum_{m\in\mathcal{P}}\sum_{s\in\mathcal{S}} c_{pm}^{\mathrm{setup}} \cdot \boldsymbol{\eta}_{pms} + \sum_{p\in\mathcal{P}}\sum_{s\in\mathcal{S}} c_p^{\mathrm{bl}} \cdot \mathbf{y}^{-}_{ps} + \sum_{p\in\mathcal{P}}\sum_{t\in\mathcal{T}} c_p^{\mathrm{sst}} \cdot \boldsymbol{\psi}_{pt}, \tag{5.14}$$

subject to

$$\mathbf{y}^{+}_{ps} - \mathbf{y}^{-}_{ps} = y_{p0} + \sum_{i\in\mathcal{S}|i\leq s} [\mathbf{q}_{pi} - (\boldsymbol{\tau}_i - \boldsymbol{\tau}_{i-1}) \cdot d_{p,t_i}], \quad \forall s \in \mathcal{S}, p \in \mathcal{P} \tag{5.15}$$

$$\sum_{p\in\mathcal{P}} k_p^{\mathrm{prd}} \cdot \mathbf{q}_{ps} + \sum_{p\in\mathcal{P}}\sum_{m\in\mathcal{P}} z_{pm}^{\mathrm{setup}} \cdot \boldsymbol{\eta}_{pms} \leq K_{t_s}^{\mathrm{prd}} \cdot (\boldsymbol{\tau}_s - \boldsymbol{\tau}_{s-1}), \quad \forall s \in \mathcal{S} \tag{5.16}$$

$$k_p^{\mathrm{inv}} \cdot (\mathbf{y}^{+}_{p,s_t^{\mathrm{last}}} - \mathbf{y}^{-}_{p,s_t^{\mathrm{last}}}) \leq K_p^{\mathrm{inv}}, \quad \forall t \in \mathcal{T}, p \in \mathcal{P} \tag{5.17}$$

$$\mathbf{y}^{+}_{p,s_t^{\mathrm{last}}} - \mathbf{y}^{-}_{p,s_t^{\mathrm{last}}} \geq y_{pt}^{\mathrm{sst}} - \boldsymbol{\psi}_{pt}, \quad \forall t \in \mathcal{T}, p \in \mathcal{P} \tag{5.18}$$

$$k_p^{\mathrm{prd}} \cdot \mathbf{q}_{ps} \leq K_{t_s}^{\mathrm{prd}} \cdot \boldsymbol{\delta}_{ps}, \quad \forall s \in \mathcal{S}, p \in \mathcal{P} \tag{5.19}$$

$$\mathbf{q}_{ps} \geq q_p^{\mathrm{min}} \cdot (\boldsymbol{\delta}_{ps} - \boldsymbol{\delta}_{ps-1}), \quad \forall s \in \mathcal{S} \setminus \mathcal{S}^{\mathrm{end}}, p \in \mathcal{P} \tag{5.20}$$

$$\mathbf{q}_{ps} + \mathbf{q}_{ps+1} \geq q_p^{\mathrm{min}} \cdot (\boldsymbol{\delta}_{ps} - \boldsymbol{\delta}_{ps-1}), \quad \forall s \in \mathcal{S}^{\mathrm{end}} \setminus |\mathcal{S}|, p \in \mathcal{P} \tag{5.21}$$

$$\boldsymbol{\tau}_s \geq \boldsymbol{\tau}_{s-1} + \frac{z^{\mathrm{minq}}}{K_{t_s}^{\mathrm{prd}}} - (1 - (\boldsymbol{\delta}_{ps} - \boldsymbol{\delta}_{p,s-1})), \quad \forall s \in \mathcal{S} \setminus \mathcal{S}^{\mathrm{end}}, p \in \mathcal{P} \tag{5.22}$$

$$(\boldsymbol{\tau}_{H_t} - \boldsymbol{\tau}_{H_t-1}) \cdot K_t^{\mathrm{prd}} + (\boldsymbol{\tau}_{H_t+1} - \boldsymbol{\tau}_{H_t}) \cdot K_{t+1}^{\mathrm{prd}} \geq$$
$$z^{\mathrm{minq}} - z^{\mathrm{minq}} \cdot (1 - (\boldsymbol{\delta}_{p,s_t^{\mathrm{last}}} - \boldsymbol{\delta}_{p,s_t^{\mathrm{last}}-1})), \quad \forall t \in \mathcal{T} \setminus |\mathcal{T}|, p \in \mathcal{P} \tag{5.23}$$

$$\sum_{p\in\mathcal{P}} \boldsymbol{\delta}_{ps} = 1, \quad \forall s \in \mathcal{S} \tag{5.24}$$

$$\boldsymbol{\delta}_{ps-1} + \boldsymbol{\delta}_{ms} \leq \boldsymbol{\eta}_{pms} + 1, \quad \forall s \in \mathcal{S}, p \in \mathcal{P}, m \in \mathcal{P} \tag{5.25}$$

$$\boldsymbol{\tau}_{s_t^{\mathrm{last}}} = t, \quad \forall t \in \mathcal{T} \tag{5.26}$$

$$\mathbf{q}_{p1} \geq q_p^{\mathrm{initial}}, \quad \forall p \in \mathcal{P} \tag{5.27}$$

$$\sum_{p\in\mathcal{P}} \boldsymbol{\eta}_{pms} = \boldsymbol{\delta}_{ms}, \quad \forall s \in \mathcal{S}, m \in \mathcal{P} \tag{5.28}$$

$$\sum_{m\in\mathcal{P}} \boldsymbol{\eta}_{pms} = \boldsymbol{\delta}_{p,s-1}, \quad \forall s \in \mathcal{S}, p \in \mathcal{P} \tag{5.29}$$

$$\boldsymbol{\delta}_{p,s-1} + \sum_{m\in\mathcal{P}|p\neq m} \boldsymbol{\eta}_{mps} + \sum_{m\in\mathcal{P}|p\neq m} (\boldsymbol{\delta}_{ms} - \sum_{l\in\mathcal{P}|l\neq m} \boldsymbol{\eta}_{lms}) \leq 1, \quad \forall s \in \mathcal{S}, p \in \mathcal{P} \tag{5.30}$$

$$\boldsymbol{\eta}_{pms} \in \{0,1\}, \quad \forall s \in \mathcal{S}, p \in \mathcal{P}, m \in \mathcal{P} \tag{5.31}$$

$$\mathbf{q}_{ps} \geq 0, \mathbf{y}^{+}_{ps} \geq 0, \mathbf{y}^{-}_{ps} \geq 0, \boldsymbol{\delta}_{ps} \in \{0,1\}, \quad \forall s \in \mathcal{S}, p \in \mathcal{P} \tag{5.32}$$

$$\boldsymbol{\psi}_{pt} \geq 0, \qquad \forall t \in \mathcal{T}, p \in \mathcal{P} \tag{5.33}$$

$$\boldsymbol{\tau}_{s} \geq 0. \qquad \forall s \in \mathcal{S} \tag{5.34}$$

The objective function (5.14) minimises the inventory holding costs, changeover costs, backlogging costs, and deviation costs from the safety stock targets. Constraints (5.15) represent the inventory balance equations on the micro period level. Constraints (5.16) ensure that the capacity requirement for the production and the changeover time in micro period s cannot exceed the available production capacity in that micro period. Constraints (5.17) ensure that the inventory level of product p at the end of macro period t (equal to s_t^{last}) will not exceed the available storage capacity for product p. The soft safety stock constraints are given in (5.18). Constraints (5.19) present the production setup logic on the micro period level. The minimum production quantity constraints are given in (5.20) and (5.21). Similarly, the minimum production time constraints are presented in (5.22) and (5.23). Since the minimum production quantities are determined based on the minimum production times, we use constraints (5.22) and (5.23) as valid inequalities. Equations (5.24) ensure that only one product can be produced in a single micro period s. Note that the equality enables the preservation of the setup state over micro periods. Constraints (5.25) ensure that changeovers are carried out at the beginning of each micro period. Equations (5.26) couple the endogenous micro periods with the exogenous macro periods. These equations impose a fixed ending time for the last micro period of every macro period t. Constraints (5.27) fix the production quantity in the first micro period to guarantee the minimum production quantity constraint related to the incomplete production from the previous planning epoch.

We further introduce valid inequalities to improve the computational efficiency. The valid inequalities (5.28) and (5.29), known as unit flow equalities, are in fact the flow conservation constraints that relate the setup and changeover variables. The valid inequalities in (5.30) take the fact into account that only four setup and changeover conditions are possible for product p in micro periods $s-1$ and s. In the first case, product p is set up in both the micro periods $s-1$ and s. In the second case, product p is set up in micro period $s-1$ but not in s. In the third case, there is a changeover for product p in micro period s, i.e., product p is set up in micro period s but not in $s-1$. Finally, in the last case, there is no setup or changeover for product p in the micro periods $s-1$ and s. For more details and proofs of these valid inequalities, we refer the reader to Koçlar (2005).

Table 5.1: Notation for the sequential approach.

Sets:	
$t \in \mathcal{T} = \{1, ..., \lvert\mathcal{T}\rvert\}$	set of macro periods
$s \in \mathcal{S} = \{1, ..., \lvert\mathcal{S}\rvert\}$	set of micro periods
$p \in \mathcal{P} = \{1, ..., \lvert\mathcal{P}\rvert\}$	set of products
$\mathcal{S}_t \subset \mathcal{S}$	set of micro periods within macro period t
$\mathcal{S}^{\text{end}} \subset \mathcal{S}$	set of last micro periods of macro periods
Parameters:	
d_{pt}	demand of product p in macro period t
q_p^{min}	minimum production quantity for product p
z_p^{minq}	minimum production time per production activity
y_{pt}^{sst}	safety stock target for product p at the end of macro period t
K_p^{inv}	storage capacity for product p
k_p^{inv}	inventory capacity consumption for a unit of product p
K_t^{prd}	length of macro period t (available production capacity)
k_p^{prd}	production capacity consumption to produce a unit of product p
z_{pm}^{setup}	changeover time from product p to product m
c_p^{h}	inventory holding cost of a unit of product p for one macro period
c_{pm}^{setup}	changeover cost from product p to product m
c_p^{bl}	contribution margin (backlog penalty cost) of a unit of product p
c_p^{sst}	cost of deviation from the safety stock level of product p per macro period
y_{p0}	initial inventory of product p at the beginning of the planning horizon
δ_{p0}	initial setup configuration of product p at the beginning of the planning horizon
q_p^{initial}	fixed production quantity which must be produced for product p in the first micro period
$\boldsymbol{\tau}_0$	the starting time of the planning horizon (= 0)
s_t^{last}	the last micro period of macro period t
t_s	the corresponding macro period of micro period s
Decision variables:	
$\mathbf{q}_{ps}$	production quantity of product p in micro period s
$\boldsymbol{\delta}_{ps}$	setup indicator of product p in micro period s

$\boldsymbol{\eta}_{pms}$	indicates if a changeover from product p to product m occurs at the beginning of micro period s
$\boldsymbol{\tau}_s$	ending time of micro period s
$\mathbf{y}^+_{ps}$	inventory on hand of product p at the end of micro period s
$\mathbf{y}^-_{ps}$	not fulfilled amount of demand (backlogged) of product p until the end of micro period s
$\boldsymbol{\psi}_{pt}$	deviation of product p from its safety stock target at the end of macro period t

5.6 Integrated Lot-sizing, Scheduling and Dynamic Safety Stock Planning Model

5.6.1 Shortage-Cost Minimisation Approach

We apply a shortage-cost minimisation approach in order to develop an integrated model. This integrated model is built on the GLSP model formulation we used in the sequential approach. However, instead of specifying the approximate length of the replenishment cycles exogenously for safety stock sizing, we integrate this task into the MILP model and determine the exact length of the replenishment cycles that result from the production schedule endogenously. Whenever we schedule the production of product p in a micro period, as well as in the very first micro period of the planning horizon, we calculate the expected backlogged quantity that results from the on-hand stock and the planned production quantity at the end of the replenishment cycle. We penalise this expected quantity in the objective function. Thus, the production quantities are chosen in such a way that they account for the demand uncertainty over the replenishment cycle.

Let $\mathfrak{b}^{\mathrm{RL}}_{ps}$ denote the expected backlogged quantities in micro period s if s is the start of a replenishment cycle, otherwise it is zero. Let $\mathbf{y}^{-\mathrm{RL}}_{ps}$ denote the deterministic backlogged quantity in every micro period s (equivalent to $\mathbf{y}^-_{ps}$) except if s is the start of a replenishment cycle. Note that, in every micro period s, only one of $\mathfrak{b}^{\mathrm{RL}}_{ps}$ and $\mathbf{y}^{-\mathrm{RL}}_{ps}$ can take on a positive value. We define $\mathbf{y}^{\mathrm{post}}_{ps} \in \mathbb{R}$ as the inventory level of product p in micro period s immediately after production and before the realisation of demand, i.e., $\mathbf{y}^{\mathrm{post}}_{ps} = \mathbf{y}^+_{p,s-1} - \mathbf{y}^-_{p,s-1} + \mathbf{q}_{ps}$. Moreover, let $f_p(x)$ and $F_p(x)$ denote the probability density and cumulative distribution functions of the demand of product p during the replenishment cycle, respectively. In the following, we present the mathematical formulation of the integrated approach. The complete additional notation requirement for the integrated approach is summarised in Table 5.2.

$$\min C = \sum_{p\in\mathcal{P}}\sum_{t\in\mathcal{T}} c^{\mathrm{h}}_p \cdot \mathbf{y}^{+}_{p,s^{\mathrm{last}}_t} + \sum_{p\in\mathcal{P}}\sum_{m\in\mathcal{P}}\sum_{s\in\mathcal{S}} c^{\mathrm{setup}}_{pm} \cdot \boldsymbol{\eta}_{pms} + \sum_{p\in\mathcal{P}}\sum_{s\in\mathcal{S}} c^{\mathrm{bl}}_p \cdot \mathfrak{b}^{\mathrm{RL}}_{ps} + \sum_{p\in\mathcal{P}}\sum_{s\in\mathcal{S}} c^{\mathrm{bl}}_p \cdot \mathbf{y}^{-\mathrm{RL}}_{ps}. \tag{5.35}$$

subject to

$$(5.15) - (5.17), (5.19) - (5.34),$$

$$\mathfrak{b}^{\mathrm{RL}}_{ps} = \int_{\mathbf{y}^{\mathrm{post}}_{ps}}^{\infty} (x - \mathbf{y}^{\mathrm{post}}_{ps}) \cdot f_p(x) \cdot dx, \qquad \forall p \in \mathcal{P}, s \in \mathcal{S} \tag{5.36}$$

$$\mathbf{y}^{-\mathrm{RL}}_{ps} \geq \mathbf{y}^{-}_{ps} - \boldsymbol{\delta}_{ps} \cdot M, \qquad \forall p \in \mathcal{P}, s \in \mathcal{S} | s > 1 \tag{5.37}$$

$$\mathbf{y}^{-\mathrm{RL}}_{ps} \geq 0. \qquad \forall p \in \mathcal{P}, s \in \mathcal{S} \tag{5.38}$$

The first and second terms in the objective function (5.35) are identical to those for the sequential approach in (5.14). We penalise the expected backlogged quantity $\mathfrak{b}^{\mathrm{RL}}_{ps}$ by the contribution margin c^{bl}_p. The last term penalises the deterministic backlog within a replenishment cycle, not including its starting micro period. This last term is only used for clearing the backlog quantities within a replenishment cycle at the earliest possible time as we penalise the expected backlog quantities only at the start of the replenishment cycles. Note that, in (5.35), we have dropped the penalty cost term for deviations from the safety stock targets ($\sum_{p\in\mathcal{P}}\sum_{t\in\mathcal{T}} c^{\mathrm{sst}}_p \cdot \boldsymbol{\psi}_{pt}$), which is no longer needed.

Constraints (5.15) - (5.17) and (5.19)-(5.34) from the sequential approach are also used in the integrated model. Note that we only drop the soft constraints for the exogenous target safety stocks (5.18). Equations (5.36) determine the expected backlogged quantity of product p at the end of the replenishment cycle that starts in micro period s. Constraints (5.37) and (5.38) determine the values of the deterministic backlogged quantity $\mathbf{y}^{-\mathrm{RL}}_{ps}$ based on $\mathbf{y}^{-}_{ps}$ in micro periods s where s is not the beginning of a replenishment cycle.

The most challenging part of the above model formulation are the non-linear equations in (5.36). We can rewrite them as

$$\mathfrak{b}_{ps}^{\mathrm{RL}} = \int_{\mathbf{y}_{ps}^{\mathrm{post}}}^{\infty} (x - \mathbf{y}_{ps}^{\mathrm{post}}) \cdot f_p(x) \cdot dx = \mathbf{b}_{ps} - \int_{0}^{\mathbf{y}_{ps}^{\mathrm{post}}} \Big[1 - F_p(x)\Big] \cdot dx, \tag{5.39}$$

where, $\mathbf{b}_{ps}$ is the mean of the replenishment cycle demand of product p in micro period s. (5.39) implies that $\mathfrak{b}_{ps}^{\mathrm{RL}}$ depends on $\mathbf{b}_{ps}$, $\mathbf{y}_{ps}^{\mathrm{post}}$ and $F_p(x)$. $F_p(x)$ further depends on the demand uncertainty parameters (mean and variance) over the replenishment cycle. In the following three sections, we focus on how to specify $\mathfrak{b}_{ps}^{\mathrm{RL}}$ step by step. We first develop an MILP to determine the mean and variance of the replenishment cycle (Section 5.6.2) endogenously. Then, in Section 5.6.3, we show how to adjust these demand uncertainty parameters in order to account for the re-planning opportunities that exist under rolling horizon planning. Finally, in Section 5.6.4, we introduce a bivariate linearisation method for determining $\mathfrak{b}_{ps}^{\mathrm{RL}}$.

5.6.2 Demand Uncertainty Parameters over Replenishment Cycles

The mean or variance of the replenishment cycle demand is defined as the sum of the mean or variance of micro periods within a replenishment cycle. Since the length of micro periods can assume non-integer values, a macro period demand is disaggregated (uniformly distributed) in the micro level proportional to the lengths of its micro periods. Mathematically, the mean or variance of micro period s belonging to macro period t for product p is written as $(\boldsymbol{\tau}_s - \boldsymbol{\tau}_{s-1}) \cdot d_{pt}$ and $(\boldsymbol{\tau}_s - \boldsymbol{\tau}_{s-1}) \cdot \sigma_{pt}^2$, respectively. Note that, here and hereafter, we assume that $\mu_{pt} = d_{pt}$ and σ_{pt}^2 indicate the mean and variance of demand of every product p in each macro period t.

- **Mean of the Replenishment Cycle Demand (MRCD)**

  We modify the approach introduced in Section 4.4.2 to take non-equidistant micro periods into account. Let $b_{pt}^{\max}$ denote the sum of the mean demand from (including) macro period t to the end of the planning horizon for product p. We introduce the decision variable $\bar{\mathbf{b}}_{ps}$ as the sum of the mean of demands of product p from micro period s to the beginning of the next production lot, expressed in the fraction of $b_{p,t_s}^{\max}$. If s is a production period, then $\bar{\mathbf{b}}_{ps}$ represents the MRCD. Moreover, let $\mathbf{u}_{ps}$ denote an auxiliary continuous decision variable corresponding to $\bar{\mathbf{b}}_{p,t_s}$. The

Table 5.2: Additional notation for the integrated approach*.

Mean of the Replenishment Cycle Demand (MRCD)	
Parameter:	
$b_{pt}^{\max}$	Upper boundary on the MRCD
Decision variables:	
$\mathbf{b}_{ps}$	MRCD
$\bar{\mathbf{b}}_{ps}$	MRCD in fraction of $b_{pt}^{\max}$
$\mathbf{u}_{ps}$	Auxiliary decision variable
Variance of the Replenishment Cycle Demand (VRCD)	
Parameter:	
$m_{pt}^{\max}$	Upper boundary on the VRCD
cov_{pij}	Covariance of product p for macro period i and j, where $j \geq i$
Decision variables:	
$\mathbf{m}_{ps}$	VRCD
$\mathbf{m}_{ps}^{\text{ind}}$	VRCD if demand is independent
$\bar{\mathbf{m}}_{ps}^{\text{ind}}$	$\mathbf{m}_{ps}^{\text{ind}}$ in fraction of $m_{pt}^{\max}$
$\mathbf{m}_{ps}^{\text{crl,micro}}$	Part of VRCD refers to the correlation between different micro periods
$\mathbf{m}_{pt}^{\text{crl}}$	Approximation of $\mathbf{m}_{ps}^{\text{crl,micro}}$ on the macro level
$\boldsymbol{\delta}_{pt}^{\text{macro}}$	Binary decision variable that takes on the value of one if there is a setup in t for p, otherwise zero.
$\mathbf{c}_{pt}^{\text{crl}}$	First auxiliary decision variable
$\mathbf{c}_{pij}^{\text{irl}}$	Second auxiliary decision variable
Expected backlog	
Decision variables:	
$\mathbf{y}_{ps}^{\text{post}}$	Inventory level after production before the realisation of demand in s for p
$\mathbf{b}_{ps}^{\text{post}}$	Defined as $\mathbf{y}_{p,s-1}^{+} - \mathbf{y}_{p,s-1}^{-} + \mathbf{q}_{ps} - \mathbf{b}_{ps}$
$\mathbf{b}_{ps}^{\text{post}+}$	Absolute value of $\mathbf{b}_{ps}^{\text{post}}$
$\mathbf{b}_{ps}^{\text{post}-}$	$-\mathbf{b}_{ps}^{\text{post}}$ if $\mathbf{b}_{ps}^{\text{post}} \leq 0$
$\boldsymbol{\gamma}_{ps}$	Binary variable that takes on value of one if $\mathbf{b}_{ps}^{\text{post}} \leq 0$, otherwise zero
$\mathfrak{b}_{ps}$	Expected backlog at the end of the replenishment cycle which includes s
$\mathfrak{b}_{ps}^{\text{RL}}$	$\mathfrak{b}_{ps}$ if s is the start of the replenishment cycle, otherwise zero
$\mathbf{y}_{ps}^{-\text{RL}}$	Deterministic backlogged quantity if s is not the start of an replenishment cycle

* s stands for micro period $s \in \mathcal{S}$, t for macro period $t \in \mathcal{T}$ and p for product $p \in \mathcal{P}$

following mathematical model formulation determines the MRCD.

$$\mathbf{u}_{ps} \leq \boldsymbol{\delta}_{ps}, \qquad \forall p \in \mathcal{P}, s \in \mathcal{S} \quad (5.40)$$

$$\mathbf{u}_{ps} \geq \bar{\mathbf{b}}_{ps} - (1 - \boldsymbol{\delta}_{ps}), \qquad \forall p \in \mathcal{P}, s \in \mathcal{S} \quad (5.41)$$

$$\mathbf{u}_{ps} \leq \bar{\mathbf{b}}_{ps} + (1 - \boldsymbol{\delta}_{ps}), \qquad \forall p \in \mathcal{P}, s \in \mathcal{S} \quad (5.42)$$

$$\bar{\mathbf{b}}_{ps} \cdot b^{\max}_{p,t_s} = \sum_{i \in S | i \geq s} (\boldsymbol{\tau}_i - \boldsymbol{\tau}_{i-1}) \cdot d_{p,t_i} - \sum_{j \in S | j > s} \mathbf{u}_{pj} \cdot b^{\max}_{p,t_j}, \qquad \forall p \in \mathcal{P}, t \in \mathcal{T} \quad (5.43)$$

$$\bar{\mathbf{b}}_{ps}, \mathbf{u}_{ps} \geq 0. \qquad \forall p \in \mathcal{P}, s \in \mathcal{S} \quad (5.44)$$

According to (5.40), if micro period s is not a production period, the auxiliary decision variable $\mathbf{u}_{ps}$ becomes zero. Constraints (5.41) and (5.42) ensure that $\mathbf{u}_{ps}$ is exactly equal to the fractional value of the cumulative mean demand ($\bar{\mathbf{b}}_{ps}$) if the micro period s is a production period. Equations (5.43) determine the MRCD for a production start in micro period s ($\mathbf{b}_{ps} = \bar{\mathbf{b}}_{ps} \cdot b^{\max}_{p,t_s}$) in a recursive way. These equations sum up the mean demand of all the following micro periods and including micro period s and subtract the sum of the auxiliary variables of micro periods j, where $j > s$ for every $j \in S$. Note that the values of the auxiliary variables in these equations are converted to the quantity-based values ($\mathbf{u}_{pj} \cdot b^{\max}_{pj}$). Constraints (5.44) indicate the non-negativity of the decision variables.

- **Variance of the Replenishment Cycle Demand (VRCD)**

  On top of the MRCD calculation, we need to introduce a new method for computing VRCD due to the serial correlation of the period demands. Let cov_{pij} denote the covariance of product p for periods i and j, where $j \geq i$. Moreover, we define cov^*_{pij} as follows.

$$cov^*_{pij} = cov_{pij}, \qquad \forall p \in \mathcal{P}, i, j \in \mathcal{T} | i = j \quad (5.45)$$

$$cov^*_{pij} = 2 \cdot cov_{pij}. \qquad \forall p \in \mathcal{P}, i, j \in \mathcal{T} | i \neq j \quad (5.46)$$

  We describe a replenishment cycle as micro periods $k \in \{i, i+1, ..., j-1\}$, starting from micro period i and ending at the beginning of micro period j. Let $\mathbf{m}_{pi}$ denote the VRCD of product p during the replenishment cycle, which starts from micro period i. The following equations specify the VRCD.

$$\mathbf{m}_{pi} = \sum_{k \in \mathcal{S} | k \geq i \wedge k < j} \; \sum_{l \in \mathcal{S} | l \geq k \wedge k < j} cov^*_{pkl}. \qquad \forall p \in \mathcal{P}, i \in \mathcal{S} \tag{5.47}$$

We can separate equations (5.47) into:

$$\mathbf{m}_{pi} = \underbrace{\sum_{k \in \mathcal{S} | k \geq i \wedge k < j} cov^*_{pkk}}_{\mathbf{m}^{\text{ind}}_{pi}} + \underbrace{\sum_{k \in \mathcal{S} | k \geq i \wedge k < j} \; \sum_{l \in \mathcal{S} | l > k \wedge l < j} cov^*_{pkl}}_{\mathbf{m}^{\text{crl,micro}}_{pi}}. \quad \forall p \in \mathcal{P}, i \in \mathcal{S} \tag{5.48}$$

The first term in (5.48), i.e., $\mathbf{m}^{\text{ind}}_{pi}$, refers to the VRCD if there is an independent demand situation, which can be determined as follows. Let $m^{\max}_{pt}$ denote the sum of the variances of demand from (including) macro period t to the end of the planning horizon for product p. We introduce the decision variable $\bar{\mathbf{m}}^{\text{ind}}_{ps}$ as the sum of the variances of product p from micro period s to the beginning of the next production lot, expressed in fraction of $b^{\max}_{p,t_s}$. $\mathbf{m}^{\text{ind}}_{pi}$ is determined analogously to the MILP formulation of the MRCD by using the following mathematical model.

$$\mathbf{c}_{ps} \leq \boldsymbol{\delta}_{ps}, \qquad \forall p \in \mathcal{P}, s \in \mathcal{S} \tag{5.49}$$

$$\mathbf{c}_{ps} \geq \bar{\mathbf{b}}_{ps} - (1 - \boldsymbol{\delta}_{ps}), \qquad \forall p \in \mathcal{P}, s \in \mathcal{S} \tag{5.50}$$

$$\mathbf{c}_{ps} \leq \bar{\mathbf{b}}_{ps} + (1 - \boldsymbol{\delta}_{ps}), \qquad \forall p \in \mathcal{P}, s \in \mathcal{S} \tag{5.51}$$

$$\bar{\mathbf{m}}^{\text{ind}}_{ps} \cdot m^{\max}_{p,t_s} = \sum_{i \in S | i \geq s} (\boldsymbol{\tau}_i - \boldsymbol{\tau}_{i-1}) \cdot \sigma^2_{p,t_i} - \sum_{j \in S | j > s} \mathbf{c}_{pj} \cdot m^{\max}_{p,t_j}, \qquad \forall p \in \mathcal{P}, t \in \mathcal{T} \tag{5.52}$$

$$\bar{\mathbf{m}}^{\text{ind}}_{ps}, \mathbf{c}_{ps} \geq 0. \qquad \forall p \in \mathcal{P}, s \in \mathcal{S} \tag{5.53}$$

The second term in (5.48), $\mathbf{m}^{\text{crl,micro}}_{pi}$, relates to the correlation between different micro periods. Assuming that the correlation coefficients are only known between macro periods, the determination of the correlation between micro periods of different lengths requires a higher sophistication level and handling effort. Thus, we use an approximation method that only calculates the corresponding covariances on the macro level. Figure 5.3 (lower part) illustrates a replenishment cycle on the micro level, starting from micro period i and ending at the beginning of micro period j. The reflection of this replenishment cycle on the macro level starts at the beginning of macro period t_i and ends at the beginning of macro period t_j. We approximate $\mathbf{m}_{pi}^{\text{crl,micro}}$ by only taking the correlation between the macro periods that are fully covered by the replenishment cycle, i.e., from t_i+1 to t_j-1 (denoted by $\mathbf{m}_{p,t_i}^{\text{crl}}$), (see upper part of Figure 5.3). Mathematically,

$$\mathbf{m}_{pi}^{\text{crl,micro}} \sim \mathbf{m}_{p,t_i}^{\text{crl}} = \sum_{t \in \mathcal{T} | t > t_i \wedge t < t_j} \sum_{k \in T | k > t \wedge k < t_j} cov_{ptk}^{*}. \quad \forall p \in \mathcal{P}, i \in \mathcal{S} \tag{5.54}$$

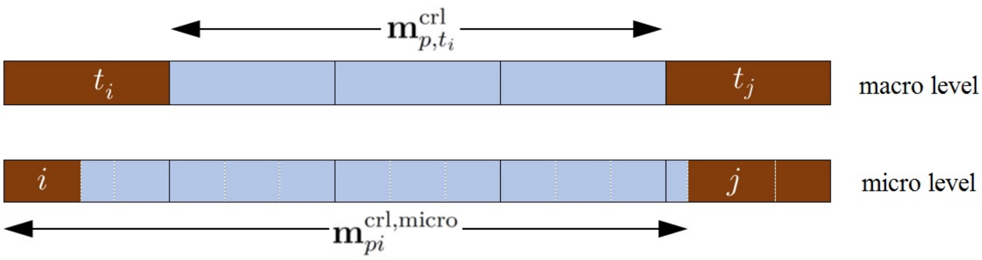


Figure 5.3: Correlation in the VRCD.

We introduce a new recursive method to determine $\mathbf{m}_{pt}^{\text{crl}}$ $\forall p \in \mathcal{P}, t \in \mathcal{T}$ on the macro level. We use the example illustrated in Figure 5.4 to explain the recursive method. Assume a planning horizon with six macro periods, where production occurs in macro periods one and four. Figure 5.4a presents all possible combinations of $cov^{*}(.)$ between these six macro periods except those taken in (5.45) into consideration. For the sake of simplicity, we drop index p. We start from production period $t = 4$, which is the starting point of the last replenishment cycle (periods four to six). According to the above explanation, $\mathbf{m}_{pt}^{\text{crl}}$ in $t = 4$ is equal to $cov_{5,6}^{*}$. We exclude $cov_{4,5}^{*}$ and $cov_{4,6}^{*}$ since we do not know if macro period 4 is fully covered by the corresponding replenishment cycle as defined on the micro period level (see Figure 5.3). We introduce a new continuous auxiliary decision variable $\mathbf{c}_{pt}^{\text{crl}}$ $\forall t \in T$ and set $\mathbf{c}_{pt}^{\text{crl}} = \mathbf{m}_{pt}^{\text{crl}}$ for $t = 4$. If t is not a production period, we set $\mathbf{m}_{pt}^{\text{crl}} = 0$. Since $t = 4$ is a production period, the combinations of $cov_{i,j}^{*}$ for $\forall i, j \in \mathcal{T} | i \leq 4 \wedge j \geq 4 \wedge i \neq j$ become irrelevant for the calculation of the correlation of replenishment cycles that precede $t = 4$ (see dark grey block in Figure 5.4a). Note that $i \leq 4$ and $j \geq 4$ with $i \neq j$ no longer belong to the same replenishment cycle. In order to capture these irrelevant combinations, we introduce another continuous auxiliary decision variable $\mathbf{c}_{pjt}^{\text{irl}}$ $\forall j, t \in \mathcal{T} | j \leq t = 4$ and set $\mathbf{c}_{pjt}^{\text{irl}} = \sum_{k \in T | k \geq t \wedge j \neq k} cov_{jk}^{*}$ for $\forall j, t \in \mathcal{T} | j \leq t$ for $t = 4$. For example, for $j = 2$ with $t = 4$, we have $\mathbf{c}_{p,2,4}^{\text{irl}} = cov_{2,4}^{*} + cov_{2,5}^{*} + cov_{2,6}^{*}$. If t is not a production period, we set $\mathbf{c}_{pjt}^{\text{irl}} = 0$ for $\forall j \in \mathcal{T} | j \leq t$.

We move back to the beginning of the replenishment cycle with the one that starts in $t = 4$, i.e., $t = 1$, where $\mathbf{m}^{\text{crl}}_{pt}$ is equal to $cov^*_{2,3}$. $\mathbf{m}^{\text{crl}}_{pt}$ with $t = 1$ (block A in Figure 5.4b) is determined as follows. We sum up all combinations of cov^*_{jk} $\forall j, k \in \mathcal{T} | j > t \wedge k > j$ (blocks A+B+C in Figure 5.4b), less $\mathbf{c}^{\text{crl}}_{pk}$ $\forall k > t$ (block C) and $\mathbf{c}^{\text{irl}}_{pij}$ $\forall i, j \in \mathcal{T} | i > t \wedge i \leq j \wedge j > t$ (block B). Both blocks C and B have already been determined in the replenishment cycle that succeeds the current one.

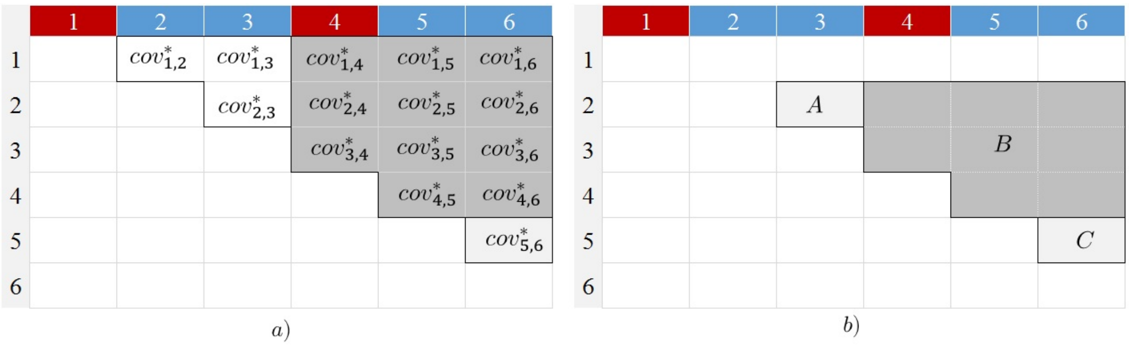


Figure 5.4: Illustration of the determination of $\mathbf{m}^{\text{crl}}_{pt}$.

The corresponding MILP is given as follows.

$$\boldsymbol{\delta}^{\text{macro}}_{p,t_s} \geq \boldsymbol{\delta}_{ps}, \qquad \forall p \in \mathcal{P}, s \in \mathcal{S} \tag{5.55}$$

$$\boldsymbol{\delta}^{\text{macro}}_{pt} \leq \sum_{s \in \mathcal{S}_t} \boldsymbol{\delta}_{ps}, \qquad \forall p \in \mathcal{P}, t \in \mathcal{T} \tag{5.56}$$

$$\boldsymbol{\delta}^{\text{macro}}_{pt} \in \{0, 1\}, \qquad \forall p \in \mathcal{P}, t \in \mathcal{T} \tag{5.57}$$

$$\mathbf{c}^{\text{crl}}_{pt} \leq \boldsymbol{\delta}^{\text{macro}}_{pt} \cdot M, \qquad \forall p \in \mathcal{P}, t \in \mathcal{T} \tag{5.58}$$

$$\mathbf{c}^{\text{crl}}_{pt} \geq -1 \cdot \boldsymbol{\delta}^{\text{macro}}_{pt} \cdot M, \qquad \forall p \in \mathcal{P}, t \in \mathcal{T} \tag{5.59}$$

$$\mathbf{c}^{\text{crl}}_{pt} \geq \mathbf{m}^{\text{crl}}_{pt} - (1 - \boldsymbol{\delta}^{\text{macro}}_{pt}) \cdot M, \qquad \forall p \in \mathcal{P}, t \in \mathcal{T} \tag{5.60}$$

$$\mathbf{c}^{\text{crl}}_{pt} \leq \mathbf{m}^{\text{crl}}_{pt} + (1 - \boldsymbol{\delta}^{\text{macro}}_{pt}) \cdot M, \qquad \forall p \in \mathcal{P}, t \in \mathcal{T} \tag{5.61}$$

$$\mathbf{c}^{\text{irl}}_{pjt} \leq \boldsymbol{\delta}^{\text{macro}}_{pt} \cdot M, \qquad \forall p \in \mathcal{P}, t \in \mathcal{T}, j \in \mathcal{T} | j \leq t \tag{5.62}$$

$$\mathbf{c}^{\text{irl}}_{pjt} \geq -1 \cdot \boldsymbol{\delta}^{\text{macro}}_{pt} \cdot M, \qquad \forall p \in \mathcal{P}, t \in \mathcal{T}, j \in \mathcal{T} | j \leq t \tag{5.63}$$

$$\mathbf{c}^{\text{irl}}_{pjt} \geq \sum_{k=t}^{|\mathcal{T}|} cov^*_{pjk} - \sum_{k=t+1}^{|\mathcal{T}|} \mathbf{c}^{\text{irl}}_{pjk} - (1 - \boldsymbol{\delta}^{\text{macro}}_{pt}) \cdot M, \qquad \forall p \in \mathcal{P}, t \in \mathcal{T}, j \in \mathcal{T} | j \leq t \tag{5.64}$$

$$\mathbf{c}^{\text{irl}}_{pjt} \leq$$

$$\sum_{k=t}^{|\mathcal{T}|} cov^*_{pjk} - \sum_{k=t+1}^{|\mathcal{T}|} \mathbf{c}^{\text{irl}}_{pjk} + (1 - \boldsymbol{\delta}^{\text{macro}}_{pt}) \cdot M, \qquad \forall p \in \mathcal{P}, t \in \mathcal{T}, j \in \mathcal{T} | j \leq t \tag{5.65}$$

$$\mathbf{m}^{\text{crl}}_{pt} = \sum_{j=t+1}^{|\mathcal{T}|} \sum_{k=j+1}^{|\mathcal{T}|} cov^*_{pjk}$$

$$- \sum_{k=t+1}^{|\mathcal{T}|} \mathbf{c}^{\text{crl}}_{pk} - \sum_{k=t+1}^{|\mathcal{T}|} \sum_{j=t+1}^{k} \mathbf{c}^{\text{irl}}_{pjk}, \qquad \forall p \in \mathcal{P}, t \in \mathcal{T} \tag{5.66}$$

$$\mathbf{m}^{\text{crl}}_{pt}, \mathbf{c}^{\text{crl}}_{pt} \in \mathbb{R}, \qquad \forall p \in \mathcal{P}, t \in \mathcal{T} \tag{5.67}$$

$$\mathbf{c}^{\text{irl}}_{pjt} \in \mathbb{R}. \qquad \forall p \in \mathcal{P}, j \in \mathcal{T} t \in \mathcal{T} | j \leq t \tag{5.68}$$

Constraints (5.55) ensure that $\boldsymbol{\delta}^{\text{macro}}_{pt}$ becomes 1 if any of the micro periods in macro period t are set up for product p. Constraints (5.56) guarantee that $\boldsymbol{\delta}^{\text{macro}}_{pt}$ becomes zero if no setup is configured for product p in any of the micro periods in macro period t. Constraints (5.57) indicate the binary decision variables. Constraints (5.58) and (5.59) set the first auxiliary variable to zero in non-production periods. Constraints (5.60) and (5.61) set the first auxiliary decision variable equal to $\mathbf{m}^{\text{crl}}_{pt}$ if t is a production period. Constraints (5.62) and (5.63) set the second auxiliary variable to zero in non-production periods. Constraints (5.64) and (5.65) assign the sum of the invalid combinations of $cov^*(.)$ in macro period t to the second auxiliary decision variable if t is a production period. Finally, constraints (5.66) determine $\mathbf{m}^{\text{crl}}_{pt}$ by summing up all the combinations of $cov^*(.)$ between the macro periods following t except those from (5.45), less the combinations that have already been considered in two auxiliary variables $\mathbf{c}^{\text{crl}}_{pk}$, $k > t$ and $\mathbf{c}^{\text{irl}}_{pjk}$, where $k > t$ and $j > t \wedge j \leq k$. The constraints in (5.67) and (5.68) indicate that the auxiliary decision variables can assume either positive or negative values, depending whether the serially-correlated demand is positive or negative.

5.6.3 Re-planning-Opportunity Adjustment under Rolling Horizon Planning

Under rolling-horizon planning, the production schedule, and thus also the previously determined replenishment cycles, can be changed at regular intervals. Changes are most likely in settings with a lot of excess (safety) capacity. We found in the previous chapter that, if we ignore this additional flexibility, the placement of dynamic safety stocks over replenishment cycles can result in excessive stocks. We can reduce dynamic safety stocks and save the corresponding costs by arranging for them to cover only part of the demand uncertainty over the replenishment cycle. The remainder is dealt with by the safety capacity and the resulting possibility to re-plan.

In the previous chapter, we introduced a re-planning opportunity coefficient α_p (see Figure 4.2 in Section 4.4.3) in order to divide uncertainty during the replenishment cycle between the safety stock and the safety capacity. We used a simple method to specify α_p according to the available average excess capacity.

In order to account for the additional flexibility, we introduce a new period-dependent scaling factor $\boldsymbol{\alpha}_{pit}$ where $\forall p \in \mathcal{P}, i \in \mathcal{T}, t \in \mathcal{T} | t \geq i$. We define the scaling factor $\boldsymbol{\alpha}_{pit}$ as the probability that the demand uncertainty for product p in period t needs to be covered by safety stock (inventory option) during the production in i where $t \geq i$. We use the example illustrated in Figure 5.5 to explain the scaling factor $\boldsymbol{\alpha}_{pit}$. Consider a production plan where periods 2 and 8 represent production periods (marked in red). When producing in period 2, we keep safety stock to cover demand uncertainty during periods 2 to 7. If there is a replanning opportunity with enough capacity flexibility, we may deploy capacity flexibility instead of keeping safety stock to cover uncertainty during all periods between 2 and 7. For example, we may already start production in period 6 instead of in period 8 if we have enough excess capacity in period 6. If that is what we do, then we only need safety stock to cover uncertainty for periods between 2 and 5. In this example, $\boldsymbol{\alpha}_{pit}$ with $i = 2$ becomes 1 for every $t = \{2, 3, 4, 5\}$ and $\boldsymbol{\alpha}_{pit} = 0$ for $t = \{6, 7\}$.

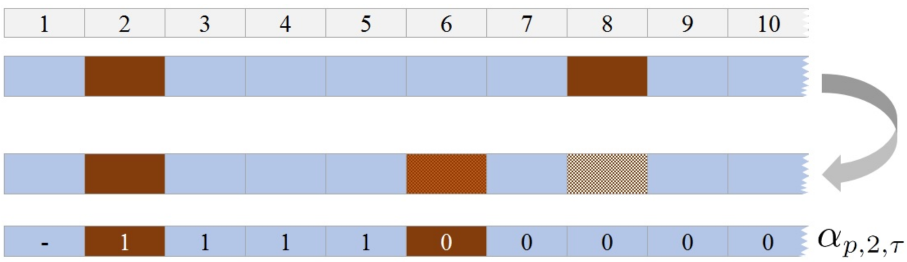


Figure 5.5: The scaling factor $\boldsymbol{\alpha}_{p,t,\tau}$ in period $t = 2$.

Being in period 2, we only need to make the advancement decision (to keep inventory or deploy capacity flexibility) for the periods prior to the next planned production period in 8, i.e., during the replenishment cycle that starts from period 2. The scaling factor is set to zero for all periods $t \geq 8$. Moreover, we must ensure that we have enough excess capacity in period 6 to cover average demand during periods 6 and 7.

In order to find the optimal values of the scaling factor $\boldsymbol{\alpha}_{pit}$, we develop the following linear program.

$$\min C = \sum_{p\in\mathcal{P},i\in\mathcal{T},t\in\mathcal{T}|t\geq i} c_p^{\mathrm{h}} \cdot z_p \cdot \sigma_{pt} \cdot \boldsymbol{\alpha}_{pit}, \tag{5.69}$$

subject to

$$\sum_{p\in\mathcal{P},k\in\mathcal{T}|k\geq t\wedge k<i+L_p} 0.5\cdot(1-\boldsymbol{\alpha}_{pit})\cdot a_p\cdot d_{pk} \leq [K_t - \sum_{p\in\mathcal{P}} a_p\cdot d_{pt}]^+, \quad \forall i\in\mathcal{T}, t\in\mathcal{T}|t\geq i \tag{5.70}$$

$$\boldsymbol{\alpha}_{pit} \geq \boldsymbol{\alpha}_{pi,t+1}, \qquad \forall p\in\mathcal{P}, i,t\in\mathcal{T}|t\geq i, t<|\mathcal{T}| \tag{5.71}$$

$$\boldsymbol{\alpha}_{pit} \leq \boldsymbol{\alpha}_{p,i+1,t}, \qquad \forall p\in\mathcal{P}, i,t\in\mathcal{T}|t\geq i+1, i<|\mathcal{T}| \tag{5.72}$$

$$0 \leq \boldsymbol{\alpha}_{pit} \leq 1. \qquad \forall p\in\mathcal{P}, i,t\in\mathcal{T}|t\geq i \tag{5.73}$$

In the objective function (5.69), we minimise the total inventory holding costs if we decide for the inventory option to cover demand uncertainty during all periods and over all products. Note that $z_p \cdot \sigma_{pt}$ is the required safety stock in period t for product p.

Constraints (5.70) ensure that we have enough excess capacity in period t if we decide in period i that the next planned production lot should be advanced to period t. On the right-hand side of constraints (5.70), the approximated excess capacity in period t is set to the total available capacity in period t (K_t) minus the total production capacity consumption for demand of all products in that period. Note that, for reasons of simplification, we assume a lot-for-lot case.

On the left-hand side of constraints (5.70), we sum up the capacity requirements of the demand from period i until $i + L_p$ for those we decide a capacity flexibility option. $i + L_p$ anticipates the start of the next planned production lot where L_p indicates the average length of a replenishment cycle. Moreover, assuming a symmetric distribution by which excess demand only occurs in 50% of the cases, we multiply the term on the left-hand side of constraints (5.70) with 0.5.

Constraints (5.71) indicate that the probability of choosing a flexibility option for the demand of a period increases if a demand is far in the future. Constraints (5.72) imply that, if a demand was assigned to the inventory option, then we do not need other re-planning opportunities, i.e., the other scaling factors that refer to the same demand in period t should also be one.

Having determined the optimal values for the scaling factor α_{pit}, we need to adjust the values of the MRCD and VRCD. In the previous chapter, we simply multiplied the re-planning opportunity coefficient α_p with the MRCD and VRCD. In the case of the scaling factor $\boldsymbol{\alpha}_{pit}$, we need to adjust the model formulations we introduced in the previous section for the MRCD and the VRCD.

We first present the adjusted model formulation for the MRCD. In every period $i \in \mathcal{T}$, the mean demand of the periods following to i, i.e., $t \in \mathcal{T}|t \geq i$, which is used for calculating the MRCD, will depend on α_{pit}. We define a matrix of demand as $D_{pit} = d_{pt} \cdot \alpha_{pit}$ where $\forall p \in \mathcal{P}, i \in \mathcal{T}, t \in \mathcal{T}|t \geq i$. Note that, $\alpha_{pit} = 1$ if $i = t$. We map this matrix on the micro level: $\mathbf{D}_{pis} = D_{pt_it_s} \cdot (\boldsymbol{\tau}_s - \boldsymbol{\tau}_{s-1})$, where t_i and t_s indicate the corresponding macro period of the micro period i and s and $(\boldsymbol{\tau}_s - \boldsymbol{\tau}_{s-1})$ is the length of micro period s. We further require two continuous auxiliary decision variables of $\mathbf{u}^{\mathrm{I}}_{pis}$ and $\mathbf{u}^{\mathrm{II}}_{pis}$ where $i, s \in \mathcal{S}|s \geq i$. The second auxiliary decision variable ($\mathbf{u}^{\mathrm{II}}_{pis}$) is used for recursively calculating the first auxiliary decision variable $\mathbf{u}^{\mathrm{I}}_{pis}$. The first auxiliary decision variable is used for calculating the MRCD. The adjusted mathematical model formulation is given as follows.

$$\mathbf{u}^{\mathrm{I}}_{pis} \leq \boldsymbol{\delta}_{ps} \cdot b^{\max}_{pt_s}, \qquad \forall p \in \mathcal{P}, i \in \mathcal{S}, s \in \mathcal{S}|s \geq i \tag{5.74}$$

$$\mathbf{u}^{\mathrm{I}}_{pis} \geq \mathbf{u}^{\mathrm{II}}_{pis} - (1 - \boldsymbol{\delta}_{ps}) \cdot b^{\max}_{pt_s}, \qquad \forall p \in \mathcal{P}, i \in \mathcal{S}, s \in \mathcal{S}|s \geq i \tag{5.75}$$

$$\mathbf{u}^{\mathrm{I}}_{pis} \leq \mathbf{u}^{\mathrm{II}}_{pis} + (1 - \boldsymbol{\delta}_{ps}) \cdot b^{\max}_{pt_s}, \qquad \forall p \in \mathcal{P}, i \in \mathcal{S}, s \in \mathcal{S}|s \geq i \tag{5.76}$$

$$\mathbf{u}^{\mathrm{II}}_{pis} = \sum_{k \in S|k \geq s} \mathbf{D}_{pik} - \sum_{k \in S|k > s} \mathbf{u}^{\mathrm{I}}_{pik}, \qquad \forall p \in \mathcal{P}, i \in \mathcal{S}, s \in \mathcal{S}|s \geq i \tag{5.77}$$

$$\mathbf{b}_{pi} = \sum_{k \in S|k \geq i} \mathbf{D}_{pik} - \sum_{k \in S|k > i} \mathbf{u}^{\mathrm{I}}_{pik}. \qquad \forall p \in \mathcal{P}, i \in \mathcal{S} \tag{5.78}$$

Constraints (5.74) set the first auxiliary decision variable $\mathbf{u}^{\mathrm{I}}_{pis}$ to zero if s is not a production period for every product p in every period i where $s \geq i$. If s is a production period, constraints (5.75) and (5.76) set $\mathbf{u}^{\mathrm{I}}_{pis}$ to the second auxiliary decision variable $\mathbf{u}^{\mathrm{II}}_{pis}$. In a recursive way similar to the one we introduced in the previous section in equations (5.43), we use the auxiliary decision variable $\mathbf{u}^{\mathrm{I}}_{pis}$ in equations (5.77) for determining the value of $\mathbf{u}^{\mathrm{II}}_{pis}$ and equations (5.78) for determining the values of $\mathbf{b}_{pi}$. Equations (5.77) and (5.78) can be seen as the period-specific versions ($i \in \mathcal{T}$) of the recursive method applied in (5.43). Note that we need equations (5.77) in order to correctly determine the values of $\mathbf{u}^{\mathrm{I}}_{pis}$ in constraints (5.75) and (5.76).

We now present the adjusted mathematical model formulation to specify $\mathbf{m}^{\text{ind}}_{pi}$. We define a matrix of variances of demand as $V_{pit} = \sigma^2_{pt} \cdot \alpha_{pit}$ where $\forall p \in \mathcal{P}, i \in \mathcal{T}, t \in \mathcal{T} | t \geq i$. We map this matrix on the micro level: $\mathbf{V}_{pis} = V_{pt_i t_s} \cdot (\boldsymbol{\tau}_s - \boldsymbol{\tau}_{s-1}) \; \forall p \in \mathcal{P}, i, s \in \mathcal{S} | s \geq i$. We define two continuous auxiliary decision variables of $\mathbf{c}^{\text{I}}_{pis}$ and $\mathbf{c}^{\text{II}}_{pis}$ where $\forall p \in \mathcal{P}, i, s \in \mathcal{S} | s \geq i$. In analogy to the adjusted model formulation for the MRCD, the following mathematical formulation determines $\mathbf{m}^{\text{ind}}_{pi} \; \forall i \in \mathcal{S}$.

$$\mathbf{c}^{\text{I}}_{pis} \leq \boldsymbol{\delta}_{ps} \cdot m^{\max}_{pt_s}, \qquad \forall p \in \mathcal{P}, i \in \mathcal{S}, s \in \mathcal{S} | s \geq i \tag{5.79}$$

$$\mathbf{c}^{\text{I}}_{pis} \geq \mathbf{c}^{\text{II}}_{pis} - (1 - \boldsymbol{\delta}_{ps}) \cdot m^{\max}_{pt_s}, \qquad \forall p \in \mathcal{P}, i \in \mathcal{S}, s \in \mathcal{S} | s \geq i \tag{5.80}$$

$$\mathbf{c}^{\text{I}}_{pis} \leq \mathbf{c}^{\text{II}}_{pis} + (1 - \boldsymbol{\delta}_{ps}) \cdot m^{\max}_{pt_s}, \qquad \forall p \in \mathcal{P}, i \in \mathcal{S}, s \in \mathcal{S} | s \geq i \tag{5.81}$$

$$\mathbf{c}^{\text{II}}_{pis} = \sum_{k \in S | k \geq s} \mathbf{V}_{pik} - \sum_{k \in S | k > s} \mathbf{c}^{\text{I}}_{pik}, \qquad \forall p \in \mathcal{P}, i \in \mathcal{S}, s \in \mathcal{S} | s \geq i \tag{5.82}$$

$$\mathbf{m}^{\text{ind}}_{pi} = \sum_{k \in S | k \geq i} \mathbf{V}_{pik} - \sum_{k \in S | k > i} \mathbf{c}^{\text{I}}_{pik}. \qquad \forall p \in \mathcal{P}, i \in \mathcal{S} \tag{5.83}$$

In the last step, we change the model formulation for $\mathbf{m}^{\text{crl}}_{pt} \; \forall p \in \mathcal{P}, t \in \mathcal{T}$ as introduced in (5.55) to (5.68). First, for every period $i \in \mathcal{T}$, we determine a period-specific $cov^*(\cdot)$ as $cov^*_{pijt} = cov^*_{pjt} \cdot \alpha_{pit}$ where $\forall p \in \mathcal{P}, j, t \in \mathcal{T} | j \geq i \wedge t \geq j$. The two auxiliary decision variables introduced in the previous section for $\mathbf{m}^{\text{crl}}_{pt}$ receive a period-specific index $\forall i \in \mathcal{T}$, i.e., $\mathbf{c}^{\text{crl}}_{pit}$ where $t \geq i$ and $\mathbf{c}^{\text{irl}}_{pijt}$ where $j \geq i$ and $t \geq j$. Moreover, we define a new auxiliary decision variable $\mathbf{u}^{\text{crl}}_{pit}$ where $t \geq i$, which is used for determining $\mathbf{c}^{\text{crl}}_{pit}$. The following mathematical model formulation determines $\mathbf{m}^{\text{crl}}_{pt}$ if we apply the scaling factor α_{pit}.

$$\mathbf{c}^{\text{crl}}_{pit} \leq \boldsymbol{\delta}^{\text{macro}}_{pt} \cdot M, \qquad \forall p \in \mathcal{P}, i, t \in \mathcal{T} | t \geq i \tag{5.84}$$

$$\mathbf{c}^{\text{crl}}_{pit} \geq -1 \cdot \boldsymbol{\delta}^{\text{macro}}_{pt} \cdot M, \qquad \forall p \in \mathcal{P}, i, t \in \mathcal{T} | t \geq i \tag{5.85}$$

$$\mathbf{c}^{\text{crl}}_{pit} \geq \mathbf{u}^{\text{crl}}_{pit} - (1 - \boldsymbol{\delta}^{\text{macro}}_{pt}) \cdot M, \qquad \forall p \in \mathcal{P}, i, t \in \mathcal{T} | t \geq i \tag{5.86}$$

$$\mathbf{c}^{\text{crl}}_{pit} \leq \mathbf{u}^{\text{crl}}_{pit} + (1 - \boldsymbol{\delta}^{\text{macro}}_{pt}) \cdot M, \qquad \forall p \in \mathcal{P}, i, t \in \mathcal{T} | t \geq i \tag{5.87}$$

$$\mathbf{c}^{\text{irl}}_{pijt} \leq \boldsymbol{\delta}^{\text{macro}}_{pt} \cdot M, \qquad \forall p \in \mathcal{P}, i, j, t \in \mathcal{T} | j \geq i \wedge t \geq j \tag{5.88}$$

$$\mathbf{c}^{\text{irl}}_{pijt} \geq -1 \cdot \boldsymbol{\delta}^{\text{macro}}_{pt} \cdot M, \qquad \forall p \in \mathcal{P}, i, j, t \in \mathcal{T} | j \geq i \wedge t \geq j \tag{5.89}$$

$$\mathbf{c}^{\text{irl}}_{pijt} \geq \sum_{k \in \mathcal{T} | k \geq t} cov^*_{pijk} - \sum_{k \in \mathcal{T} | k > t} \mathbf{c}^{\text{irl}}_{pijk} - (1 - \boldsymbol{\delta}^{\text{macro}}_{pt}) \cdot M, \quad \forall p \in \mathcal{P}, i, j, t \in \mathcal{T} | j \geq i \wedge t \geq j \tag{5.90}$$

$$\mathbf{c}^{\text{irl}}_{pijt} \leq \sum_{k\in\mathcal{T}|k\geq t} cov^*_{pijk} - \sum_{k\in\mathcal{T}|k>t} \mathbf{c}^{\text{irl}}_{pijk} + (1-\boldsymbol{\delta}^{\text{macro}}_{pt}) \cdot M, \quad \forall p \in \mathcal{P}, i,j,t \in \mathcal{T}|j \geq i \wedge t \geq j \quad (5.91)$$

$$\mathbf{u}_{pit} = \sum_{j\in\mathcal{T}|j>t}\sum_{k\in\mathcal{T}|k\geq j} cov^*_{pijk} - \sum_{k\in\mathcal{T}|k>t} \mathbf{c}^{\text{crl}}_{pik} - \sum_{k\in\mathcal{T}|k>t}\sum_{j\in\mathcal{T}|j>t\wedge j\leq k} \mathbf{c}^{\text{irl}}_{pijk}, \quad \forall p \in \mathcal{P}, i,t \in \mathcal{T}|t \geq i \quad (5.92)$$

$$\mathbf{m}^{\text{crl}}_{pt} = \sum_{j\in\mathcal{T}|j>t}\sum_{k\in\mathcal{T}|k\geq j} cov^*_{ptjk} - \sum_{k\in\mathcal{T}|k>t} \mathbf{c}^{\text{crl}}_{ptk} - \sum_{k\in\mathcal{T}|k>t}\sum_{j\in\mathcal{T}|j>t\wedge j\leq k} \mathbf{c}^{\text{irl}}_{ptjk}, \quad \forall p \in \mathcal{P}, t \in \mathcal{T} \quad (5.93)$$

$$\mathbf{m}^{\text{crl}}_{pt}, \quad \forall p \in \mathcal{P}, t \in \mathcal{T} \quad (5.94)$$

$$\mathbf{c}^{\text{crl}}_{pit}, \mathbf{u}^{\text{crl}}_{pit} \in \mathbb{R}, \quad \forall p \in \mathcal{P}, i,t \in \mathcal{T}|t \geq i \quad (5.95)$$

$$\mathbf{c}^{\text{irl}}_{pijt} \in \mathbb{R}. \quad \forall p \in \mathcal{P}, i,j,t \in \mathcal{T}|j \geq i \wedge t \geq j \quad (5.96)$$

The constraints in (5.84)- (5.96) are the period-specific versions of the original model formulation introduced in (5.55) to (5.68) with an extra index of $\forall i \in \mathcal{T}$. Note that, in the period-specific version, we further added an intermediate step in constraints (5.92) to determine the new auxiliary decision variable $\mathbf{u}^{\text{II}}_{pit}$ $\forall p \in \mathcal{P}, i,t \in \mathcal{T}|s \geq i$. In constraints (5.86) and (5.87), we use $\mathbf{u}^{\text{II}}_{pit}$ to specify $\mathbf{c}^{\text{crl}}_{pit}$.

5.6.4 Linearisation of Expected Backlogged Quantity

Given that there is $\mathbf{b}_{ps}$ and $\mathbf{m}_{ps}$ in every micro period s and every product p, we can rewrite the equation in (5.36) as follows.

$$\forall p \in \mathcal{P}, s \in \mathcal{S}:$$

$$\mathfrak{b}_{ps}(\mathbf{b}_{ps}, \mathbf{m}_{ps}, \mathbf{y}^{\text{post}}_{ps}) = \int_{\mathbf{y}^{\text{post}}_{ps}}^{\infty} (x - \mathbf{y}^{\text{post}}_{ps}) \cdot f_p(x) \cdot dx = \mathbf{b}_{ps} - \int_0^{\mathbf{y}^{\text{post}}_{ps}} \Big[1 - F_p(x)\Big] \cdot dx, \quad (5.97)$$

where, $\mathfrak{b}_{ps}$ presents the expected backlogged quantity of product p in every micro period s. Note that, we temporarily use $\mathfrak{b}_{ps}$ instead of $\mathfrak{b}^{\text{RL}}_{ps}$. We will derive $\mathfrak{b}^{\text{RL}}_{ps}$ from $\mathfrak{b}_{ps}$ later on.

The expected backlogged quantity depends on three decision variables $\mathbf{b}_{ps}$, $\mathbf{m}_{ps}$ and $\mathbf{y}^{\text{post}}_{ps}$. Therefore, we face a three-dimensional non-linear function. In what follows, we show how this can be reduced to a two-dimensional non-linear function.

In the case of a normally-distributed demand for product p, we can work with the standard normal distribution instead of the normal distribution by converting $\mathbf{y}_{ps}^{\text{post}}$ into $\mathbf{z}_{ps}$ according to $\mathbf{z}_{ps} = (\mathbf{y}_{p,s-1}^{+} - \mathbf{y}_{p,s-1}^{-} + \mathbf{q}_{ps} - \mathbf{b}_{ps})/\sqrt{\mathbf{m}_{ps}}$. Consequently, the expected backlogged quantity of product p that we take in micro period s into account can be written as follows:

$$\mathfrak{b}_{ps} = \sqrt{\mathbf{m}_{ps}} \cdot \int_{\mathbf{z}_{ps}}^{\infty} (x - \mathbf{z}_{ps}) \cdot \phi_p(x) \cdot dx = \sqrt{\mathbf{m}_{ps}} \cdot \left[\int_{\mathbf{z}_{ps}}^{\infty} x \cdot \phi_p(x) \cdot dx - \mathbf{z}_{ps} \cdot \int_{\mathbf{z}_{ps}}^{\infty} \phi_p(x) \cdot dx \right]. \tag{5.98}$$

Based on the property of the standard normal distribution, that $\frac{\phi(x)}{dx} = -x \cdot \phi(x)$ holds true, the above expression is reduced to:

$$\mathfrak{b}_{ps} = \sqrt{\mathbf{m}_{ps}} \cdot \Big[\phi_p(\mathbf{z}_{ps}) - \mathbf{z}_{ps} \cdot (1 - \Phi_p(\mathbf{z}_{ps})) \Big]. \tag{5.99}$$

Let $\mathbf{b}_{ps}^{\text{post}} = \mathbf{y}_{p,s-1}^{+} - \mathbf{y}_{p,s-1}^{-} + \mathbf{q}_{ps} - \mathbf{b}_{ps}$ with $\mathbf{b}_{ps}^{\text{post}} \in \mathbb{R}$. Thus we can write $\mathbf{z}_{ps} = \mathbf{b}_{ps}^{\text{post}}/\sqrt{\mathbf{m}_{ps}}$. Consequently, the expected backlogged quantity function becomes a bivariate non-linear function, $\mathfrak{b}_{ps}(\mathbf{m}_{ps}, \mathbf{b}_{ps}^{\text{post}})$. In $\mathfrak{b}_{ps}(\mathbf{m}_{ps}, \mathbf{b}_{ps}^{\text{post}})$, for every $t \in \mathcal{T}$ and every $s \in \mathcal{S}_t$, we have $\mathbf{m}_{ps} \in [0, \sum_{k \in T | k \geq t} \sum_{l \in T | l \geq k} cov_{pkl}]$ and $\mathbf{b}_{ps}^{\text{post}} \in [y_{p0} - \sum_{i \in S | i \leq s} (\boldsymbol{\tau}_i - \boldsymbol{\tau}_{i-1}) \cdot d_{p,t_i}, b_{ps}^{\text{post,max}}]$, where $b_{ps}^{\text{post,max}}$ is the maximum inventory level of product p in micro period s. We observe that $\mathbf{b}_{ps}^{\text{post}}$ can become negative. If we let $\mathbf{b}_{ps}^{\text{post+}}$ denote the absolute value of $\mathbf{b}_{ps}^{\text{post}}$, then the following equations are always true.

$$\mathfrak{b}_{ps}(\mathbf{m}_{ps}, -\mathbf{b}_{ps}^{\text{post+}}) = \mathbf{b}_{ps}^{\text{post+}} + \mathfrak{b}_{ps}(\mathbf{m}_{ps}, \mathbf{b}_{ps}^{\text{post+}}). \tag{5.100}$$

Proof of equation (5.100)

$$\mathfrak{b}_{ps}(\mathbf{m}_{ps}, -\mathbf{b}_{ps}^{\text{post+}}) = \sqrt{\mathbf{m}_{ps}} \cdot \Big[\phi_p(\frac{-\mathbf{b}_{ps}^{\text{post+}}}{\sqrt{\mathbf{m}_{ps}}}) - \frac{-\mathbf{b}_{ps}^{\text{post+}}}{\sqrt{\mathbf{m}_{ps}}} (1 - \Phi_p(\frac{-\mathbf{b}_{ps}^{\text{post+}}}{\sqrt{\mathbf{m}_{ps}}})) \Big],$$

according to $\phi(-x) = \phi(x)$, we can write

$$= \sqrt{\mathbf{m}_{ps}} \cdot \phi_p(\frac{\mathbf{b}_{ps}^{\text{post+}}}{\sqrt{\mathbf{m}_{ps}}}) + \mathbf{b}_{ps}^{\text{post+}} - \mathbf{b}_{ps}^{\text{post+}} \Phi_p(\frac{-\mathbf{b}_{ps}^{\text{post+}}}{\sqrt{\mathbf{m}_{ps}}})$$
$$= \mathbf{b}_{ps}^{\text{post+}} + \sqrt{\mathbf{m}_{ps}} \cdot \Big[\phi_p(\frac{\mathbf{b}_{ps}^{\text{post+}}}{\sqrt{\mathbf{m}_{ps}}}) - \frac{\mathbf{b}_{ps}^{\text{post+}}}{\sqrt{\mathbf{m}_{ps}}} \Phi_p(\frac{-\mathbf{b}_{ps}^{\text{post+}}}{\sqrt{\mathbf{m}_{ps}}}) \Big].$$

As we know $\Phi(-x) = 1 - \Phi(x)$, we can rewrite the above expression as

$$= \mathbf{b}_{ps}^{\text{post+}} + \sqrt{\mathbf{m}_{ps}} \cdot \Big[\phi_p(\frac{\mathbf{b}_{ps}^{\text{post+}}}{\sqrt{\mathbf{m}_{ps}}}) - \frac{\mathbf{b}_{ps}^{\text{post+}}}{\sqrt{\mathbf{m}_{ps}}}(1 - \Phi_p(\frac{\mathbf{b}_{ps}^{\text{post+}}}{\sqrt{\mathbf{m}_{ps}}}))\Big]$$
$$= \mathbf{b}_{ps}^{\text{post+}} + \mathfrak{b}_{ps}(\mathbf{m}_{ps}, \mathbf{b}_{ps}^{\text{post+}}).$$

□

Let $\mathbf{b}_{ps}^{\text{post}-}$ denote the term $-\mathbf{b}_{ps}^{\text{post}}$ if $\mathbf{b}_{ps}^{\text{post}} \leq 0$, and 0 otherwise. Furthermore, we define a new binary decision variable $\boldsymbol{\gamma}_{ps}$, which takes on the value 1 if $\mathbf{b}_{ps}^{\text{post}} \leq 0$, and 0 otherwise. Then, the following MILP determines the values of $\mathbf{b}_{ps}^{\text{post+}}$ and $\mathbf{b}_{ps}^{\text{post}-}$.

$$\mathbf{b}_{ps}^{\text{post+}} \geq \mathbf{b}_{ps}^{\text{post}}, \qquad \forall p \in \mathcal{P}, s \in \mathcal{S} \quad (5.101)$$
$$\mathbf{b}_{ps}^{\text{post+}} \leq \mathbf{b}_{ps}^{\text{post}} + \boldsymbol{\gamma}_{ps} \cdot M, \qquad \forall p \in \mathcal{P}, s \in \mathcal{S} \quad (5.102)$$
$$\mathbf{b}_{ps}^{\text{post+}} \geq -\mathbf{b}_{ps}^{\text{post}}, \qquad \forall p \in \mathcal{P}, s \in \mathcal{S} \quad (5.103)$$
$$\mathbf{b}_{ps}^{\text{post+}} \leq -\mathbf{b}_{ps}^{\text{post}} + (1 - \boldsymbol{\gamma}_{ps}) \cdot M, \qquad \forall p \in \mathcal{P}, s \in \mathcal{S} \quad (5.104)$$
$$\mathbf{b}_{ps}^{\text{post}-} \geq \mathbf{b}_{ps}^{\text{post+}} - (1 - \boldsymbol{\gamma}_{ps}) \cdot M, \qquad \forall p \in \mathcal{P}, s \in \mathcal{S} \quad (5.105)$$
$$\boldsymbol{\gamma}_{ps} \in \{0,1\}, \mathbf{b}_{ps}^{\text{post}} \in \mathbb{R}, \mathbf{b}_{ps}^{\text{post+}} \geq 0, \mathbf{b}_{ps}^{\text{post}-} \geq 0. \qquad \forall p \in \mathcal{P}, s \in \mathcal{S} \quad (5.106)$$

In these formulations, big$M$ can be derived based on the maximum value of $\mathbf{b}_{ps}^{\text{post}}$.

In (5.100), we approximate $\mathfrak{b}_{ps}(\mathbf{m}_{ps}, \mathbf{b}_{ps}^{\text{post+}})$ by applying a bivariate linearisation technique. Before describing the linearisation technique, we first show, how to derive the values of $\mathfrak{b}_{ps}^{\text{RL}}$ from $\mathfrak{b}_{ps}$ in the following MILP. $\mathfrak{b}_{ps}^{\text{RL}}$ takes on the value of $\mathfrak{b}_{ps}$ if s is the starting micro period of the replenishment cycle. Note that a replenishment cycle starts with either $s = 1$ or a production period.

$$\mathfrak{b}_{ps}^{\text{RL}}(\mathbf{m}_{ps}, \mathbf{b}_{ps}^{\text{post+}}) \geq \mathbf{b}_{ps}^{\text{post}-} + \mathfrak{b}_{ps}(\mathbf{m}_{ps}, \mathbf{b}_{ps}^{\text{post+}}), \qquad \forall p \in \mathcal{P}, s = 1 \quad (5.107)$$
$$\mathfrak{b}_{ps}^{\text{RL}}(\mathbf{m}_{ps}, \mathbf{b}_{ps}^{\text{post+}}) \geq \mathbf{b}_{ps}^{\text{post}-} + \mathfrak{b}_{ps}(\mathbf{m}_{ps}, \mathbf{b}_{ps}^{\text{post+}}) - (1 - \boldsymbol{\delta}_{ps}) \cdot M, \qquad \forall p \in \mathcal{P}, s \in \mathcal{S} \quad (5.108)$$
$$\mathfrak{b}_{ps}(\mathbf{m}_{ps}, \mathbf{b}_{ps}^{\text{post+}}) \geq 0. \qquad \forall p \in \mathcal{P}, s \in \mathcal{S} \quad (5.109)$$

Constraints (5.107) and (5.108) ensure that the value of $\mathfrak{b}_{ps}^{\text{RL}}$ is at least equal to the sum of $\mathfrak{b}_{ps}(\mathbf{m}_{ps}, \mathbf{b}_{ps}^{\text{post+}})$ and $\mathbf{b}_{ps}^{\text{post}-}$, if s is the first micro period or a production period, respectively. The big$M$ in this formulation is determined through the maximum value that $\mathbf{b}_{ps}^{\text{post}-} + \mathfrak{b}_{ps}(\mathbf{m}_{ps}, \mathbf{b}_{ps}^{\text{post+}})$ can assume. Constraints (5.109) indicate non-negativity.

In what follows, we briefly introduce a bivariate linearisation technique based on a triangulation method which is a similar approach to the one has been used in the previous chapter at the end of Section 4.4.2. However, in the current problem, we will improve upon the approximation grid in a different way. We define $|\mathcal{N}|$ predetermined approximation points on the x-axis and $|\mathcal{M}|$ predetermined approximation points on the y-axis. For every combination of the predetermined approximation points on the x and y axes (the small circle points), we calculate the corresponding expected backlog. Furthermore, every rectangle on the grid is divided into upper and lower triangles which are associated with a binary decision variable. Depending on the triangle in which the combination of $\mathbf{b}_{ps}^{\text{post+}}$ and $\mathbf{m}_{ps}$ is located, the corresponding value of the expected backlog is interpolated by the associated weights to the vertices of that triangle.

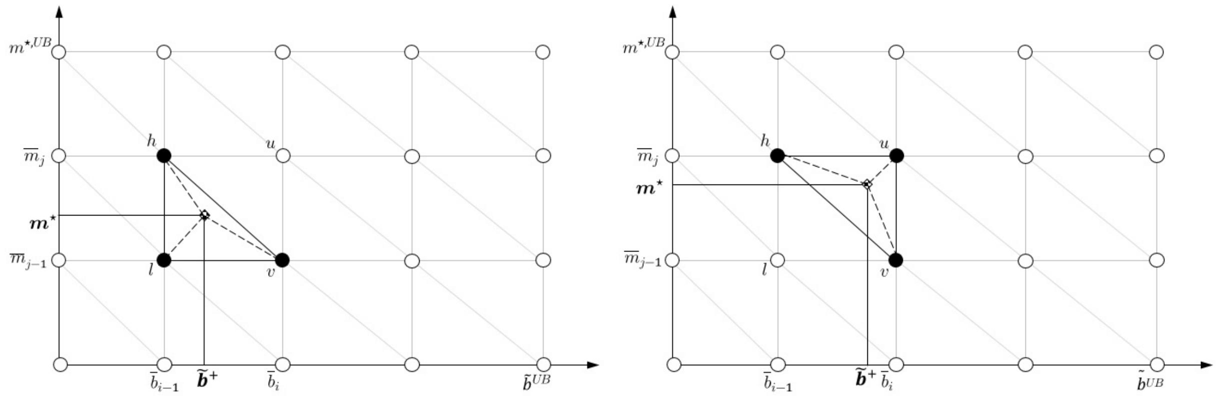


Figure 5.6: Approximation grid to linearise the expected backlog function based on the triangulation method.

We introduce a new way of improving the performance of the approximation grid by reducing the number of binary variables. In every micro period s and for each product p, we calculate an upper bound for $\mathbf{m}_{ps}$ as follows.

$$m_{ps}^{\max} = \sum_{i \in \mathcal{T} | i \geq t_s} \sum_{j \in \mathcal{T} | j \geq t_s} cov_{pij}. \qquad \forall p \in \mathcal{P}, s \in \mathcal{S} \tag{5.110}$$

Based on $m_{ps}^{\max}$, we can find a value of $b_{ps}^{\text{post,mid}}$, by which $\Phi(z_{ps}) \rightarrow 1$, where $z_{ps} = b_{ps}^{\text{post,mid}} / m_{ps}^{\max}$. Based on these definitions, we can prove that the following expression is always true for every combination of $\mathbf{m}_{ps}$, $\mathbf{b}_{ps}^{\text{post+}}$.

$$\text{if } \mathbf{m}_{ps} \leq m_{ps}^{\max} \text{ and } \mathbf{b}_{ps}^{\text{post+}} \geq b_{ps}^{\text{post,mid}} \qquad \Rightarrow \qquad \mathfrak{b}_{ps}^{\text{RL}}(\mathbf{m}_{ps}, \mathbf{b}_{ps}^{\text{post+}}) \rightarrow 0. \tag{5.111}$$

Proof of equations (5.111)

We know that $\Phi(z_{ps})$ is a non-decreasing function where $\Phi(+\infty) = 1$. Moreover, if $\Phi(z_{ps}) \to 1$, then $\phi(z_{ps}) \to 0$. According to definition, $\Phi(z_{ps}) \to 1$ with $z_{ps} = b_{ps}^{\text{post,mid}}/m_{ps}^{\max}$. For every $\mathbf{m}_{ps} \leq m_{ps}^{\max}$ and $\mathbf{b}_{ps}^{\text{post+}} \geq b_{ps}^{\text{post,mid}}$, we have $\mathbf{z}_{ps} = \frac{\mathbf{b}_{ps}^{\text{post+}}}{\sqrt{\mathbf{m}_{ps}}} \geq z_{ps}$. Thus, if $\Phi(z_{ps}) \to 1$, then $\Phi(\mathbf{z}_{ps}) \to 1$ and consequently $\phi(\mathbf{z}_{ps}) \to 0$. Recalling the equations in (5.99), for every combination of $\mathbf{m}_{ps}$ and $\mathbf{b}_{ps}^{\text{post+}}$ where $\mathbf{m}_{ps} \leq m_{ps}^{\max}$ and $\mathbf{b}_{ps}^{\text{post+}} \geq b_{ps}^{\text{post,mid}}$, we have:

$$\mathfrak{b}_{ps}^{\text{RL}}(\mathbf{m}_{ps}, \mathbf{b}_{ps}^{\text{post+}}) = \sqrt{\mathbf{m}_{ps}} \cdot \Big[\phi_p(\mathbf{z}_{ps}) - \mathbf{z}_{ps} \cdot (1 - \Phi_p(\mathbf{z}_{ps}))\Big] \to 0.$$

□

Therefore, if $b_{ps}^{\text{post,max}} > b_{ps}^{\text{post,mid}}$, it is sufficient to introduce only two triangles (two binary decision variables) for the whole area where $\mathbf{b}_{ps}^{\text{post+}}$ ranges between $b_{ps}^{\text{post,mid}}$ and $b_{ps}^{\text{post,max}}$ (see Figure 5.7).

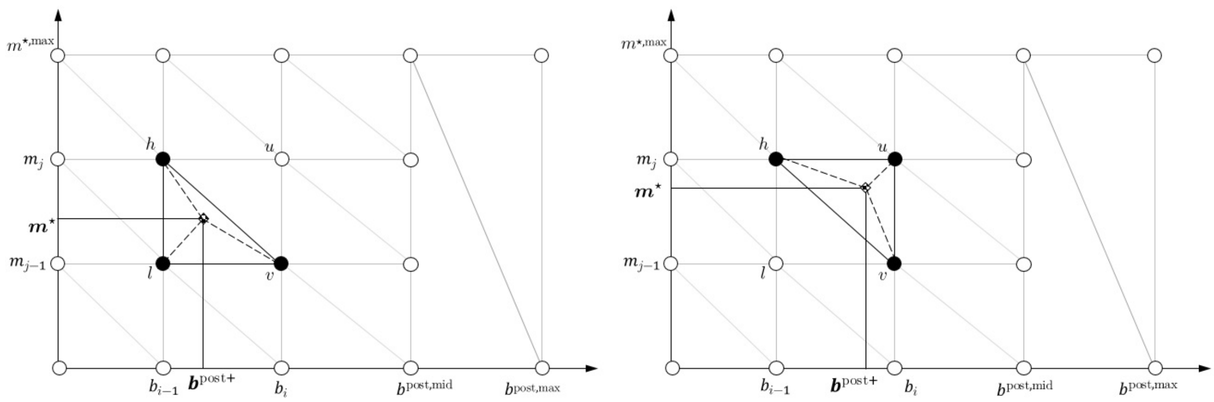


Figure 5.7: The improved approximation grid.

In the following, we present the mathematical formulation of the approximation grid. The notation is given in Table 5.3.

$\forall p \in \mathcal{P}, s \in \mathcal{S}$:

$$\sum_{i\in\hat{\mathcal{N}}}\sum_{j\in\hat{\mathcal{M}}} \boldsymbol{\omega}_{psij}^{\text{u}} + \sum_{i\in\bar{\mathcal{N}}}\sum_{j\in\bar{\mathcal{M}}} \boldsymbol{\omega}_{psij}^{\text{l}} + \boldsymbol{\omega}_{ps}^{\text{U}} + \boldsymbol{\omega}_{ps}^{\text{L}} = 1, \tag{5.112}$$

$$\boldsymbol{\lambda}_{ps}^{\text{u}'} + \boldsymbol{\lambda}_{ps}^{\text{u}'\text{v}'} + \boldsymbol{\lambda}_{ps}^{\text{u}'\text{h}'} = \boldsymbol{\omega}_{ps}^{\text{U}}, \tag{5.113}$$

$$\boldsymbol{\lambda}_{ps}^{\text{l}'} + \boldsymbol{\lambda}_{ps}^{\text{l}'\text{v}'} + \boldsymbol{\lambda}_{ps}^{\text{l}'\text{h}'} = \boldsymbol{\omega}_{ps}^{\text{L}}, \tag{5.114}$$

$$\boldsymbol{\lambda}^{\mathrm{u}}_{psij} + \boldsymbol{\lambda}^{\mathrm{uv}}_{psij} + \boldsymbol{\lambda}^{\mathrm{uh}}_{psij} = \boldsymbol{\omega}^{\mathrm{u}}_{psij}, \qquad i \in \hat{\mathcal{N}}, j \in \hat{\mathcal{M}} \quad (5.115)$$

$$\boldsymbol{\lambda}^{\mathrm{l}}_{psij} + \boldsymbol{\lambda}^{\mathrm{lv}}_{psij} + \boldsymbol{\lambda}^{\mathrm{lh}}_{psij} = \boldsymbol{\omega}^{\mathrm{l}}_{psij}, \qquad i \in \bar{\mathcal{N}}, j \in \bar{\mathcal{M}} \quad (5.116)$$

$$\begin{aligned}
\mathbf{b}^{\mathrm{post+}}_{ps} = {} & \\
& \sum_{i\in\hat{\mathcal{N}}}\sum_{j\in\hat{\mathcal{M}}} [(\boldsymbol{\lambda}^{\mathrm{u}}_{psij} + \boldsymbol{\lambda}^{\mathrm{uv}}_{psij}) \cdot b_{psi} + \boldsymbol{\lambda}^{\mathrm{uh}}_{psij} \cdot b_{ps,i-1}] + \sum_{i\in\bar{\mathcal{N}}}\sum_{j\in\bar{\mathcal{M}}} [(\boldsymbol{\lambda}^{\mathrm{l}}_{psij} + \boldsymbol{\lambda}^{\mathrm{lv}}_{psij}) \cdot b_{psi} + \boldsymbol{\lambda}^{\mathrm{lh}}_{psij} \cdot b_{ps,i+1}] \\
& + [(\boldsymbol{\lambda}^{\mathrm{u}'}_{ps} + \boldsymbol{\lambda}^{\mathrm{u}'\mathrm{v}'}_{ps}) \cdot b^{\mathrm{post,max}}_{ps} + \boldsymbol{\lambda}^{\mathrm{u}'\mathrm{h}'}_{ps} \cdot b^{\mathrm{post,mid}}_{ps}] + [(\boldsymbol{\lambda}^{\mathrm{l}'}_{ps} + \boldsymbol{\lambda}^{\mathrm{l}'\mathrm{v}'}_{ps}) \cdot b^{\mathrm{post,mid}}_{ps} + \boldsymbol{\lambda}^{\mathrm{l}'\mathrm{h}'}_{ps} \cdot b^{\mathrm{post,max}}_{ps}],
\end{aligned} \quad (5.117)$$

$$\begin{aligned}
\mathbf{m}_{ps} = {} & \sum_{i\in\hat{\mathcal{N}}}\sum_{j\in\hat{\mathcal{M}}} [(\boldsymbol{\lambda}^{\mathrm{u}}_{psij} + \boldsymbol{\lambda}^{\mathrm{uh}}_{psij}) \cdot m_{psj} + \boldsymbol{\lambda}^{\mathrm{uv}}_{psij} \cdot m_{ps,j-1}] \\
& + \sum_{i\in\bar{\mathcal{N}}}\sum_{j\in\bar{\mathcal{M}}} [(\boldsymbol{\lambda}^{\mathrm{l}}_{psij} + \boldsymbol{\lambda}^{\mathrm{lh}}_{psij}) \cdot m_{psj} + \boldsymbol{\lambda}^{\mathrm{lv}}_{psij} \cdot m_{ps,j+1}] + (\boldsymbol{\lambda}^{\mathrm{u}'}_{ps} + \boldsymbol{\lambda}^{\mathrm{u}'\mathrm{h}'}_{ps}) \cdot m^{\mathrm{max}}_{ps} + \boldsymbol{\lambda}^{\mathrm{l}'\mathrm{v}'}_{ps} \cdot m^{\mathrm{max}}_{ps},
\end{aligned} \quad (5.118)$$

$$\begin{aligned}
\mathfrak{b}^{\mathrm{RL}}_{ps} = {} & \sum_{i\in\hat{\mathcal{N}}}\sum_{j\in\hat{\mathcal{M}}} (\boldsymbol{\lambda}^{\mathrm{u}}_{psij} e^{\mathrm{backlog}}_{pij} + \boldsymbol{\lambda}^{\mathrm{uh}}_{psij} e^{\mathrm{backlog}}_{p,i-1,j} + \boldsymbol{\lambda}^{\mathrm{uv}}_{psij} e^{\mathrm{backlog}}_{p,i,j-1}) \\
& + \sum_{i\in\bar{\mathcal{N}}}\sum_{j\in\bar{\mathcal{M}}} (\boldsymbol{\lambda}^{\mathrm{l}}_{psij} e^{\mathrm{backlog}}_{pij} + \boldsymbol{\lambda}^{\mathrm{lh}}_{psij} e^{\mathrm{backlog}}_{p,i+1,j} + \boldsymbol{\lambda}^{\mathrm{lv}}_{psij} e^{\mathrm{backlog}}_{p,i,j+1}),
\end{aligned} \quad (5.119)$$

$$\boldsymbol{\omega}^{\mathrm{u}}_{psij} \in \{0,1\}, \qquad i \in \hat{\mathcal{N}}, j \in \hat{\mathcal{M}} \quad (5.120)$$

$$\boldsymbol{\omega}^{\mathrm{l}}_{psij} \in \{0,1\}, \qquad i \in \bar{\mathcal{N}}, j \in \bar{\mathcal{M}} \quad (5.121)$$

$$\boldsymbol{\omega}^{\mathrm{U}}_{ps} \in \{0,1\}, \quad (5.122)$$

$$\boldsymbol{\omega}^{\mathrm{L}}_{ps} \in \{0,1\}, \quad (5.123)$$

$$\boldsymbol{\lambda}^{\mathrm{u}}_{psij}, \boldsymbol{\lambda}^{\mathrm{uh}}_{psij}, \boldsymbol{\lambda}^{\mathrm{uv}}_{psij} \geq 0, \qquad i \in \hat{\mathcal{N}}, j \in \hat{\mathcal{M}} \quad (5.124)$$

$$\boldsymbol{\lambda}^{\mathrm{l}}_{psij}, \boldsymbol{\lambda}^{\mathrm{lh}}_{psij}, \boldsymbol{\lambda}^{\mathrm{lv}}_{psij} \geq 0, \qquad i \in \bar{\mathcal{N}}, j \in \bar{\mathcal{M}} \quad (5.125)$$

$$\boldsymbol{\lambda}^{\mathrm{u}'}_{ps}, \boldsymbol{\lambda}^{\mathrm{u}'\mathrm{h}'}_{ps}, \boldsymbol{\lambda}^{\mathrm{u}'\mathrm{v}'}_{ps} \geq 0, \quad (5.126)$$

$$\boldsymbol{\lambda}^{\mathrm{l}'}_{ps}, \boldsymbol{\lambda}^{\mathrm{l}'\mathrm{h}'}_{ps}, \boldsymbol{\lambda}^{\mathrm{l}'\mathrm{v}'}_{ps} \geq 0. \quad (5.127)$$

Constraints (5.112) ensure that, only one triangle can be used for the linearisation for each product and micro period. Constraints (5.113)-(5.116) guarantee that only the weights associated with the vertices of the selected triangle from (5.112) can obtain positive values. Consequently, the weights associated with the vertices of the other triangles become zero. Constraints (5.117) and (5.118) assign weights to the vertices of the selected triangle to build values of $\mathbf{b}_{ps}^{\text{post+}}$ and $\mathbf{m}_{ps}$, with respect to their corresponding predetermined approximation points associated with the vertices of the selected triangle. Finally, in constraints (5.119), the expected backlog is interpolated according to the determined weights of the vertices of the selected triangle and the predetermined backlog values associated with the vertices of that triangle. The non-negativity and binary decision variables are given in constraints (5.120) to (5.127).

Table 5.3: Notation used for the approximation grid.

Sets:	
$i \in \hat{\mathcal{N}} = \{1, ..., \lvert\mathcal{N}\rvert - 1\}$	set of the predetermined approximation points on x-axis, corresponding to the upper triangles
$j \in \hat{\mathcal{M}} = \{1, ..., \lvert\mathcal{M}\rvert - 1\}$	set of the predetermined approximation points on y-axis, corresponding to the upper triangles
$i \in \bar{\mathcal{N}} = \{0, ..., \lvert\mathcal{N}\rvert - 2\}$	set of the predetermined approximation points on x-axis, corresponding to the lower triangles
$j \in \bar{\mathcal{M}} = \{0, ..., \lvert\mathcal{M}\rvert - 2\}$	set of the predetermined approximation points on y-axis, corresponding to the lower triangles
Parameters:	
b_{psi}	predetermined approximation point i on the x-axis
m_{psj}	predetermined approximation point j on the y-axis
$e_{psij}^{\text{backlog}}$	predetermined expected backlog for the combination of b_{psi} and m_{psj}
Decision variables:	
$\boldsymbol{\lambda}_{psij}^{\text{u}}, \boldsymbol{\lambda}_{psij}^{\text{uv}}, \boldsymbol{\lambda}_{psij}^{\text{uh}}$	weights associated with the vertices of upper triangles
$\boldsymbol{\lambda}_{psij}^{\text{l}}, \boldsymbol{\lambda}_{psij}^{\text{lv}}, \boldsymbol{\lambda}_{psij}^{\text{lh}}$	weights associated with the vertices of lower triangles
$\boldsymbol{\lambda}_{ps}^{\text{u}'}, \boldsymbol{\lambda}_{ps}^{\text{u}'\text{v}'}, \boldsymbol{\lambda}_{ps}^{\text{u}'\text{h}'}$	weights associated with the vertices of the big upper triangles
$\boldsymbol{\lambda}_{ps}^{\text{l}'}, \boldsymbol{\lambda}_{ps}^{\text{l}'\text{v}'}, \boldsymbol{\lambda}_{ps}^{\text{l}'\text{h}'}$	weights associated with the vertices of the big lower triangles

$\boldsymbol{\omega}^{\mathrm{u}}_{psij}$	binary decision variable which gets one if the approximation point is located in the upper triangle ij, otherwise zero.
$\boldsymbol{\omega}^{\mathrm{l}}_{psij}$	binary decision variable which gets one if the approximation point is located in the lower triangle ij, otherwise zero.
$\boldsymbol{\omega}^{\mathrm{U}}_{ps}$	binary decision variable which gets one if the approximation point is located in the big upper triangle, otherwise zero.
$\boldsymbol{\omega}^{\mathrm{L}}_{ps}$	binary decision variable which gets one if the approximation point is located in the big lower triangle, otherwise zero.

5.7 Solution Approaches

5.7.1 Period-based Decomposition Heuristic

The decomposition approach divides the problem into several sub-problems that are then solved iteratively. In each sub-problem, only a reduced amount of binary decision variables is considered. The solutions of the iterations are then reassembled to from an overall solution.

We design our period-based decomposition heuristic as follows. We divide the planning horizon into decision windows, approximation windows and frozen windows (see Figure 5.8). Within a decision window, all variables and constraints of the problem are taken into account in order to find an optimal solution. A part of the solved decision window becomes a frozen window for the following iterations. In the frozen window, we fix the setup variables ($\boldsymbol{\delta}_{ps}$) based on the results of the earlier iterations. The decision windows are followed by an approximation window. In the relaxation window, we relax the setup binary decision variables, the minimum production quantity constraints and the minimum production time constraints.

We initially set the lengths of the windows for the first iteration as follows. The length of the frozen window is set to zero and the length of the decision window is set to L^{DW}. Consequently, the length of the approximation window becomes $|\mathcal{T}| - 0 - L^{\mathrm{DW}}$. For the following iterations, we increase the length of the frozen window by Δ^{FW} and decrease the length of the approximation window by Δ^{AW}. Consequently, the length of the decision window is changed by $\Delta^{\mathrm{DW}} = \Delta^{\mathrm{AW}} - \Delta^{\mathrm{FW}}$. The procedure of reducing the length of the approximation window is repeated until the last period of the decision window is the end of the planning horizon. Thus, the planning horizon of the last iteration only includes the frozen window and the decision window, as illustrated in Figure 5.8. The final solution is the result of solving the last iteration.

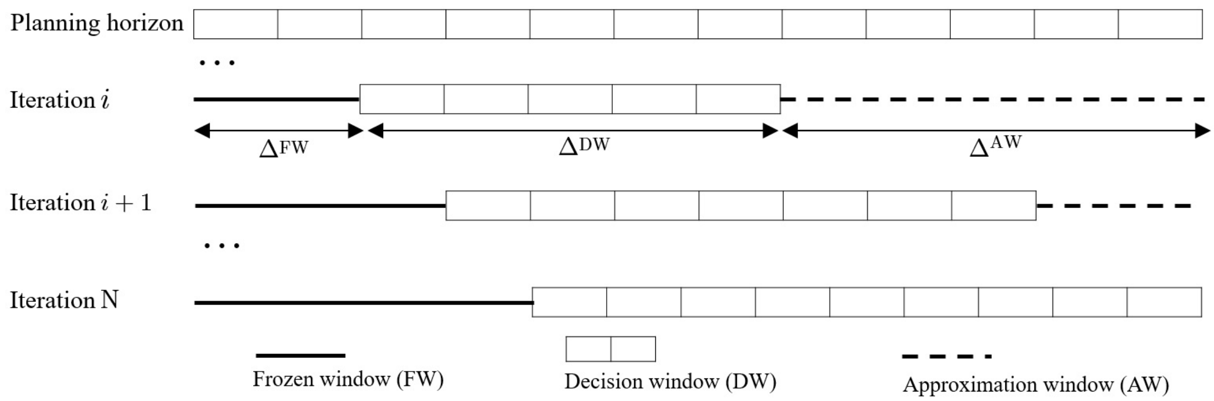


Figure 5.8: Period-based decomposition heuristic.

5.8 Numerical Analysis

5.8.1 Data

We base our numerical analysis on the real-world data set described in Section 5.2. Due to confidentiality reasons we cannot provide a more detailed data description. In addition to the production, cost and demand information, we also received the historical forecast time series that has been created each month for the next 12 months between January 2011 and December 2015. Due to structural changes in the time series, only data from January 2012 onwards could be used for products 3 and 6.

For safety stock sizing, we need an estimate of the forecast error for each product as explained in Section 5.4. We divide the provided demand and forecast time series into an estimation phase (Products 1, 2, 4 and 5: Jan 2011 - Dec 2013; Products 3 and 6: Jan 2012 - Dec 2013) and an evaluation phase (Jan 2014 - Dec 2015).

Apart from the forecast mean and standard deviation, we also need to know the distribution of the forecast for safety stock sizing in our models. With the help of the Kolmogorov-Smirnov test and the Chi-squared test at a significance level of 5%, we can confirm the fit of the normal distribution for all products that are forecasted.

In addition to the original real-world dataset, we consider three variants of it in order to analyse the sensitivity of our findings. First, we reduce the setup costs by half (Setup0.5). In the original dataset they make up the largest chunk of the total cost. Note that the setup times remain unchanged. Second, we decrease the minimum production time from 14 days to seven days (minL0.5). Consequently, the minimum production quantities are also reduced by half. Third, we reduce the production capacity consumption coefficient by 20% for all products in order to obtain a moderately capacitated situation (Capa0.8).

5.8.2 Evaluation Approach

The sequential and integrated modelling approaches presented in Sections 5.5 and 5.6 are decision generators for prescribing lot-sizes and safety stocks (Schneeweiss, 2003). These approaches use the forecast data available at the point in time when planning takes place. The actual evaluation of the generated production plans must be done based on the resulting setup, holding and backlog penalty costs under realised demands. We compute the inventory holding costs at the end of each macro period and the backlog cost at the end of each micro period where we penalise the newly backlogged quantity. The 24 months of the years 2014 and 2015 represent the evaluation horizon.

With respect to the safety stock target specification, we treat product 6 slightly different from the other products. As outlined in Section 5.2, we know exactly in which periods a positive demand occurs for this product. Consequently, we set the safety stock target to zero in all periods with zero demand to avoid an unnecessary inventory build-up.

For the original real-world dataset, as well as for the three variants, we compare the performance of the following approaches with each other, which is listed below in increasing order of their level of planning sophistication:

1. Sequential Approach (SA) with exogenous safety stocks determined according to the RoT;
2. SA with exogenous cost-optimised safety stocks (iDoS, cDoS);
3. Integrated Model (IM);
4. Integrated Model with serial demand CoRReLation (IM-CRRL).

In the IM we only calculate the endogenous safety stocks over the replenishment cycles for the products the production of which is scheduled to start in the first three macro periods. For the rest of the macro periods we use exogenous safety stocks determined according to the cost-based iDoS logic. In the considered rolling horizon planning approach with a re-planning interval of one month, only the production plan of the first macro period is implemented in every planning iteration. The endogenous determination of dynamic safety stocks in every planning iteration for the periods closer to the end of the planning horizon would unnecessarily increase the computational complexity without having a major impact on the final results.

We solve all models with the MIP solver FICO Xpress Optimizer 64-Bit v.28.01.04 on a platform with sufficient RAM and the CPU specification: Intel Core i7-4770 CPU @ 3.40 GHz, 64-bit.

5.8.3 Results

Period-based decomposition heuristic

We compare the computational performance of solving the Full Mixed-Integer Linear Programs (F-MILPs) for the SA and for the IM with a period-based Decomposition Heuristic (DH). In the SA, we use the cost-optimised cDoS safety stock targets. We solve all models for the original dataset. In addition, we solve the SA for the other three dataset variants.

For each model, we solve 24 planning iterations that correspond to our evaluation horizon under rolling horizon planning with a re-planning interval of one month. We set the maximum runtime to six hours per planning iteration for the F-MILP. In the case of the DH, the maximum runtime for solving each sub-problem is one hour. To ensure a fair comparison, we assume the same values for the initial inventory, the initial setup and the fixed production quantity for both solution approaches in every planning iteration.

Table 5.4 presents the average runtimes and their standard deviations, as well as the average optimality gaps and their standard deviations. It also shows the number of iterations (in %) where both approaches return identical objective function values (F-MILP=DH) or the DH results in a lower value (DH>F-MILP). The last two columns present, if a solution approach outperforms the other one (>), the mean and standard deviation of the differences in the objective function values, given in the fraction of the objective function value returned by the F-MILP in %.

From the SA results, we observe that solving the original datasets, Setup0.5 and Capa0.8, is considerably easier than variant minL0.5. Average runtimes and gaps are much lower. For these three settings, the DH obtains solutions comparable to those of the F-MILP, but after a much shorter runtime. On average, it takes the DH less than an hour, whereas the F-MILP needs more than three hours. For setting minL0.5, we find that the DH outperforms the F-MILP with respect to both the runtime and the solution quality (lower objective function value) in most of the iterations.

For the IM, the F-MILP returns relatively high average optimality gaps after the maximum runtime. The DH, on the other hand, returns significantly better solutions (on average up to 14% lower objective function values) after considerably shorter runtimes.

In summary, we find that the DH is a promising solution approach for this kind of model, which is why we use it for the numerical studies in the following sections.

Benefit of more sophisticated planning approaches.

We assess the potential benefit of an increase in the sophistication level of the planning approaches specified in Section 5.8.2 as levels 1-4. In particular, we want to find answers to the two questions raised by our industry partner:

1. How well does the current RoT for the safety stock determination in the SA perform compared to a cost-based safety stock optimisation?

2. What is the cost-saving potential of using an even more sophisticated approach in the form of the integrated (stochastic) model?

First, we analyse the original real-world dataset. The first block in Table 5.5 summarises the total costs and its individual components on the different planning levels. With respect to the first question, we observe a significant cost-saving potential if the safety stocks are determined with the help of the cost-optimisation approach instead of the current RoT, i.e., if we move from level 1 to level 2. The cost-optimised cDoS target results in 7% lower costs than the RoT. With the iDoS targets, 4% in costs can be saved. Concerning the second question, we find that an increase of the planning sophistication level from 2 to 3 towards our integrated model shows another considerable improvement potential. It amounts to 10% more than the cDoS target and 12% more than the iDoS targets on level 2. If we move from level 3 to level 4, i.e., IM-CRRL, we observe another 2% cost reduction.

By looking at the individual cost components, we see that the more sophisticated planning approaches tend to better solve the trade-off between the different cost types. The relative importance of the setup costs as a part of the total costs increases. The same is true for the inventory holding costs, whereas the share of the backlog costs clearly decreases. The ability to better balance the different costs results in an improved overall performance.

Table 5.4: Comparison of F-MILP with DH.

		F-MILP		DH		Comparison		Diff. of obj. function values	
Approach	Dataset	Runtime avg[stDev] (hrs.)	gap avg[stDev] (%)	Runtime avg[stDev] (hrs.)	gap avg[stDev] (%)	DH=F-MILP (%)	DH>F-MILP (%)	DH>F-MILP avg[stDev] (%)	DH<F-MILP avg[stDev] (%)
SA cDoS	original	3.5 [2.6]	7.9 [12.6]	0.6 [0.3]	0.5 [1.3]	75.0	8.3	0.3 [0.1]	2.4 [1.4]
SA cDoS	Setup0.5	3.6 [2.4]	6.1 [8.4]	0.2 [0.3]	0.0 [0.1]	54.2	4.2	2.2 [0.0]	7.9 [15.7]
SA cDoS	Capa0.8	4.4 [2.2]	9.7 [9.8]	0.4 [0.4]	0.2 [0.7]	20.8	4.2	0.0 [0.0]	1.3 [1.4]
SA cDoS	minL0.5	6.0 [0.0]	30.0 [5.2]	1.8 [1]	2.9 [3.2]	12.5	50.0	1.6 [1.2]	1.9 [1.6]
IM	original	5.4 [1.4]	25.8 [21.0]	0.9 [0.8]	0.0 [0.0]	20.8	70.8	6.4 [7.0]	0.9 [0.8]

Based on these findings, we conclude for the original dataset that it is beneficial to increase the planning level from 1 to 2. This transition can be easily done as it does not require any adjustment of the current optimisation model in APS. Only the safety stock targets for the individual products need to be determined according to the cost-optimisation logic outlined in Section 5.5.2. Consequently, the required effort is small, but cost savings of up to 7% can be achieved, depending on the DoS approach. Even though a transition from level 2 to 3 would require the replacement of the current optimisation model with the integrated one and thus some additional implementation effort, the return is significant with cost savings of at least another 10%. Moreover, another advantage of the IM is that the planner no longer has to regularly update the safety stock targets, because the model takes care of safety stock sizing endogenously. A further increase of the planning level from level 3 to 4 that leads to a cost-saving potential of 2% requires additional model complexity caused by the consideration of the serial demand correlation. The supposedly higher initial implementation effort also needs to be seen in the light of the effort savings to be expected over the subsequent years.

Next, we study the results of the three variants of the original dataset as a form of sensitivity analysis of the above findings. (see Table 5.5, blocks 2-3). For setting minL0.5, we observe that the total costs on all planning levels are higher than the ones in the original dataset, even though the same solutions should be feasible under minL0.5. The reason is that, due to the shorter minimum production quantities, the problem gets harder to solve. After our maximum computation time, the solver still has not found a solution that is at least as good as the one of the original setting. While the average gap of the decomposition heuristic for the other variants is close to 0%, it amounts to 3% for the setting minL0.5 (see Table 5.4). Since we are only interested in the relative performance of the approaches on the different planning levels and not necessarily in the absolute cost values and the performance across different datasets, we can still use these results for our analysis.

Across the three variants, we identify a total cost improvement of up to 10% if we move from level 1 to 2 either with the cDoS or with the iDoS target. This range is very similar to the original dataset. In setting Setup0.5, we observe a slight cost increase of 1% for both the cDoS and iDoS targets. This difference is insignificant if we recall that our evaluation is only based on 24 months. A total cost difference of about 1% is less than cost of one setup in our dataset. It is easily possible that, an extra setup is scheduled in the last period of the evaluation interval (period 24) just to satisfy the demand beyond the

Table 5.5: Cost overview of the different planning approaches.

Dataset	Le-el	Planning approach	Safety stock	Total costs (TC)	Rel. cost diff. to previous level	Setup costs	Share of TC	Inventory holding costs	Share of TC	Backlog costs	Share of TC
Original	1	SA	RoT	2,290,523		963,600	42%	557,832	24%	769,091	34%
	2	SA	cDoS	2,137,062	7%	924,000	43%	588,772	28%	624,290	29%
			iDoS	2,193,624	4%	924,000	42%	568,501	26%	701,124	32%
	3	IM		1,933,477	10% (vs. cDoS)	994,400	51%	585,381	30%	353,696	18%
					12% (vs. iDoS)						
	4	IM-CRRL		1,898,858	2% (vs. IM)	1,003,200	53%	590,704	31%	304,954	16%
minL0.5	1	SA	RoT	2,346,082		1,227,600	52%	485,103	21%	633,379	27%
	2	SA	cDoS	2,108,732	10%	1,098,900	52%	526,423	25%	483,409	23%
			iDoS	2,171,732	7%	1,064,800	49%	547,788	25%	559,144	26%
	3	IM		2,062,846	2% (vs. cDoS)	1,151,700	56%	553,184	27%	357,962	17%
					5% (vs. iDoS)						
	4	IM-CRRL		2,013,352	2% (vs. IM)	1,230,900	61%	587,102	29%	195,350	10%
Setup0.5	1	SA	RoT	1,705,250		462,000	27%	578,144	34%	665,107	39%
	2	SA	cDoS	1,726,798	-1%	462,000	27%	592,633	34%	672,165	39%
			iDoS	1,724,885	-1%	521,400	30%	601,476	35%	602,010	35%
	3	IM		1,383,884	20% (vs. cDoS)	501,600	36%	580,772	42%	301,511	22%
					20% (vs. iDoS)						
	4	IM-CRRL		1,396,475	-1% (vs. IM)	501,600	36%	593,167	42%	301,708	22%
Capa0.8	1	SA	RoT	1,769,617		963,600	54%	615,368	35%	190,649	11%
	2	SA	cDoS	1,645,374	7%	963,600	59%	579,463	35%	102,311	6%
			iDoS	1,594,347	10%	907,500	57%	584,458	37%	102,389	6%
	3	IM		1,809,871	-10% (vs. cDoS)	1,059,300	59%	604,936	33%	145,635	8%
					-14% (vs. iDoS)						
	4	IM-CRRL		1,719,372	5% (vs. IM)	1,019,700	59%	647,137	38%	52,535	3%

evaluation interval, e.g., in period 25. Consequently, we can conclude that the transition to the cDoS and iDoS targets on level 2 generally leads to a cost improvement.

Between the cDoS and iDoS targets, we do not observe significant differences. In setting minL0.5, the cDoS target leads to a cost-saving of 10% and the iDoS to a cost-saving of 7%. In setting Setup0.5, the performance is almost the same. In setting Capa0.8, the cDoS target provides a cost-saving of 7% and the iDoS one of 10%.

The use of the integrated model on level 3 shows a clear additional benefit for settings with a high capacity utilisation such as minL0.5 and Setup0.5. Here, the additional cost-saving potential compared to the cDoS is 2% and 20%, respectively. Compared to the iDoS target, cost savings of 5% and 20% can be achieved. However, in setting Capa0.8 with a lower capacity utilisation, we observe an increase in cost by 10% and 14% depending on the DoS approaches. In such settings, the regular re-planning possibility provides additional flexibility in the presence of sufficient capacity, which is limited in other settings with high capacity utilisation. This is an obvious shortcoming of the integrated approach that neglects this flexibility and places more safety stocks than necessary in setting Capa0.8.

If we move from level 3 to 4, we observe a total cost improvement of 2% and 5% for variants minL0.5 and Capa0.8. In Setup0.5, we observe a slight cost increase of 1%. This difference is again less than one setup cost and can be explained by the finite evaluation interval. Thus, we can conclude that the integration of the serial demand correlation into the IM leads to cost-savings of up to 5%.

With the exception of Capa0.8, we find a consistent cost improvement potential over the sequential approaches on level 1 and 2 from our newly developed integrated models. The results also confirm, as we have already observed in the original dataset, that the more sophisticated approaches choose a different balance between various cost components and thus can realise a better overall cost performance. The setup and inventory holding cost percentage of the total costs tend to increase whereas the backlog cost percentage decreases.

With respect to the poor performance of the integrated approaches in setting Capa0.8, we explore the effectiveness of our proposed re-planning opportunity scaling factor in the following section.

Benefit of re-planning opportunity scaling factor

Recall that we suggested to introduce the re-planning opportunity scaling factor, α_{pit} $\forall p \in \mathcal{P}, i, t \in \mathcal{T} | t \geq i$, to account for the additional re-scheduling flexibility under rolling horizon planning. We assume that, in settings with considerable excess capacity, the original production plan will be modified more frequently, which results in shorter actual replenishment cycles and thus a smaller safety stock requirement.

Table 5.6 summarises the lengths of the first replenishment cycles in 2014 for all products. The column "expected" shows the results that our integrated model on level 3 prescribes. The column "realised" shows the lengths of the replenishment cycles of the final production plan under rolling horizon planning.

For the original dataset, which is characterised by a high capacity utilisation of about 96%, we observe that the difference is only minor. On average across all products, the relative deviation is only 2%. In contrast, for the dataset Capa0.8 with an average capacity utilisation of 81%, we find that the average relative difference is 13%; this means the replenishment cycle length of the final production plan is shorter by 13% than the one that the integrated approach works with for the safety stock sizing.

With the help of our new scaling factor α_{pit}, the average relative difference decreases to about 7%. This illustration demonstrates that our re-planning opportunity scaling factor shows the intended effect.

Table 5.6: Replenishment cycle length comparison for first replenishment cycle in 2014.

	Original			Capa0.8					
				without α_{pit}			with α_{pit}		
Product	expected	realised	Rel. diff.	expected	realised	Rel. diff.	expected	realised	Rel. diff.
1	4.38	4.77	-9%	5.47	4.16	24%	4.16	4.22	-1%
2	5.40	7.14	-32%	7.20	7.55	-5%	8.68	8.32	4%
3	12.00	11.34	6%	12.00	10.68	11%	12.00	10.74	11%
4	3.13	2.90	7%	3.80	3.34	12%	3.61	3.20	11%
5	1.42	1.13	20%	2.29	1.00	56%	1.52	1.25	18%
6	2.06	1.73	16%	1.80	1.50	16%	2.04	2.00	2%
Avg.	4.73	4.84	-2%	5.42	4.70	13%	5.34	4.95	7%

In order to assess the economic benefit of the re-planning opportunity scaling factor, we compute the total costs of the integrated models with and without α_{pit}.

Figure 5.9 shows that the cost-saving potential due to α_{pit} amounts to about 7.9% for the IM on level 3 and is more than 1.9% for the IM-CRRL. These results demonstrate that, even though our way of determining α_{pit} is heuristic, it can improve the performance of the integrated approaches.

To sum up, both the DoS target approaches and the newly developed integrated models represent approaches that produce consistent results across the analysed parameter settings and can therefore be strongly recommended for implementation in practice. Even though both DoS targets can be integrated more easily into the existing APS than the

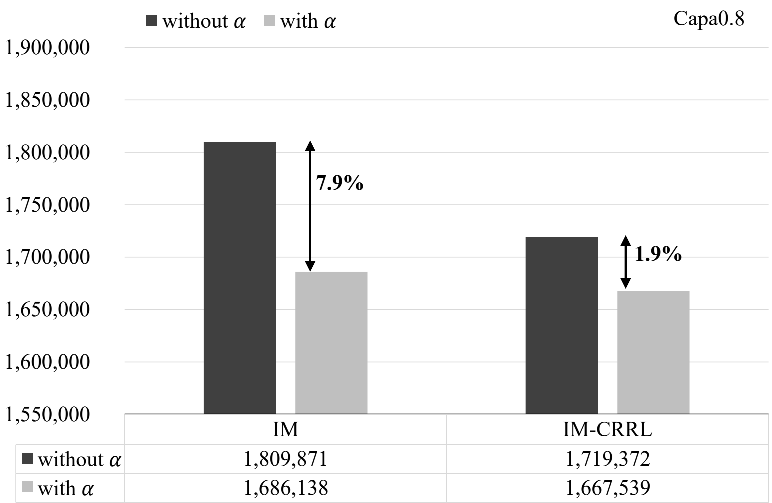


Figure 5.9: Cost-saving potential due to re-planning opportunity adjustment.

IM, a single target DoS value is also more easily manageable, which is another aspect that further advocates the use of the cDoS target in practice.

Figure 5.10 graphically illustrates the magnitude of the cost-saving potential of different approaches compared to the widely-used RoT at our industrial partner. While the transition from level 1 to level 2 enables mainly some quick wins in the form of moderate cost savings, a further transition to the integrated model on level 3 (IM) and level 4 (IM-CRRL) usually unlocks a much larger cost-saving potential.

5.9 Conclusions

We studied a simultaneous lot-sizing and scheduling problem with stochastic demands. In the literature, as well as in most APS, this optimisation problem is commonly translated into a deterministic problem by the introduction of predetermined safety stocks to the lot-sizing and scheduling model formulation. We developed a new MILP that endogenously determines dynamic safety stocks for both uncorrelated and serially-correlated demands. To account for the re-scheduling opportunities under rolling horizon planning, we introduced a heuristic that prevents an excessive build-up of safety stocks. We used this new model formulation to help our industrial partner with the quantification of his cost-saving potential that results from the use of more sophisticated production planning approaches.

Based on a real-world dataset, we found that a cost-based optimisation of the exogenous safety stock targets in the traditional deterministic APS optimisation model can deliver considerable value compared to a widely-used rule-of-thumb at our industrial partner.

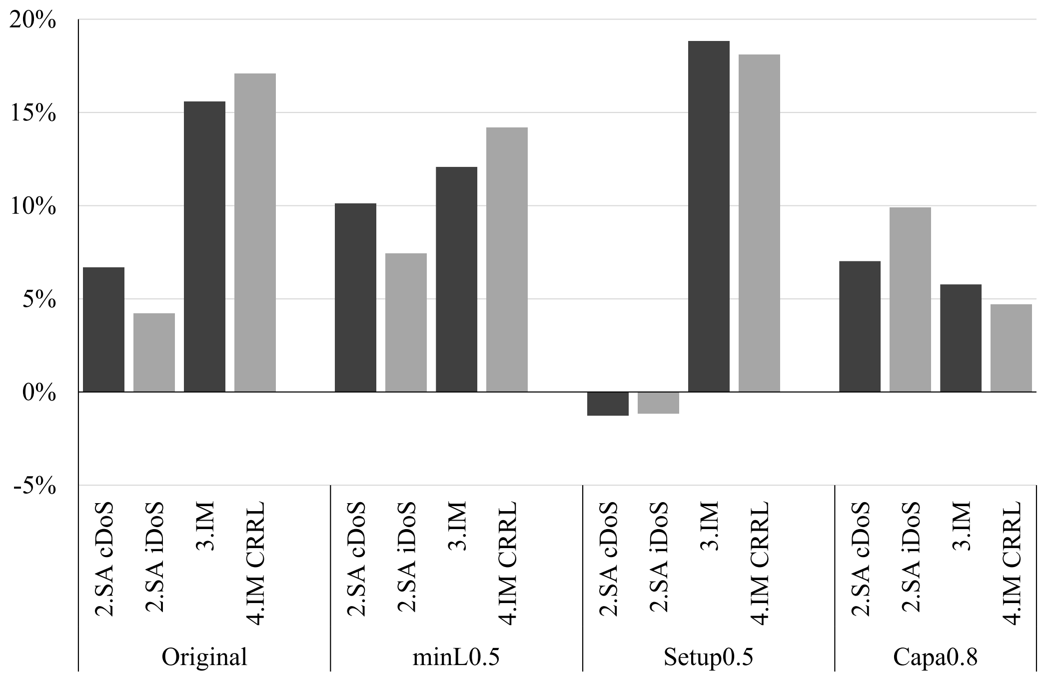


Figure 5.10: Cost-saving potential of different planning approaches compared to the rule-of-thumb.

Both DoS approaches can return cost-savings of up to 10%. While the performance of both the iDoS and cDoS targets are almost similar, a single cost-optimised cDoS target for all products is easily implementable in practice. The newly developed integrated models, on the other hand, prove robust across all studied scenarios. This makes it very appealing for application in practice. Compared to the cDoS or iDoS results, the integrated approaches provide an additional improvement potential of up to 20%. Despite its complexity, the integration of the serial demand correlation into the IM, i.e., the use of the IM-CRRL, can return cost-savings of up to 5%. Another appealing feature of the integrated approaches is that the planner does not have to worry about the appropriate specification of the replenishment cycle for the safety stock determination at all, as this is all done endogenously.

Chapter 6

Conclusions

6.1 Summary

In the context of APP, this thesis first addressed a capacity planning problem in a case where customer demand showed significant fluctuations. This problem was faced by a highly automated electronics manufacturer with multi-stage production lines and parallel machines located at different facilities. In order to determine the optimal capacity levels, three flexibility instruments were available: Shift planning, overtime account and flexible maintenance. This thesis introduced an MILP to set capacity levels at different workstations while a cost-optimal trade-off between flexibility instruments, subcontracting and inventories was obtained. An extensive numerical analysis validated the robustness of the proposed approach and its computational efficiency by using either a commercial or an open-source MIP-solver. The cost analysis of using flexibility instruments showed significant cost-saving potentials resulting from shift planning. It was further revealed that using a combination of different flexibility instruments had a higher cost-saving potential than the sum of the cost-saving obtained from each individual flexibility instrument. The simultaneous application of various types of flexibility instruments was especially important in the case of limited capacity, where obtaining higher flexibility was crucial for the adjustment of capacity levels at workstations in response to demand fluctuations.

Following a hierarchical planning concept and moving from APP to more disaggregate planning, this thesis further considered the S-CLSP and S-GLSP.

An optimal solution of the S-CLSP required a simultaneous determination of dynamic safety stocks and lot-sizes in the presence of limited capacity and stochastic demand. For this problem, dynamic safety stocks were required to meet a target customer service level in terms of fill-rate. This stochastic problem was described by means of an SDP. Due to the curse of dimensionality of the SDP, three MILPs were proposed as alternative modelling approaches, namely, (i) a sequential approach, which sequentially solved the safety stock planning problem and the deterministic CLSP with predetermined safety stocks, (ii) an integrated model based on chance-constrained programming and (iii) an adjusted integrated model to include re-planning opportunities under rolling horizon planning. While the integrated model endogenously determined dynamic safety stocks over non-equidistant replenishment cycles, the adjusted integrated model further reduced these endogenous dynamic safety stocks according to the available safety capacities. In both variants of the integrated model, a bivariate linearisation technique based on a triangulation method was applied for the linear interpolation of the non-linear order-up-to-level function. The linearisation technique was further improved by excluding all unnecessary binary decision variables on the approximation grid. The computational tests revealed a significant improvement in the performance of the bivariate linearisation technique through the use of the adjusted approximation grid.

In an extensive experimental study, the performances of the proposed modelling approaches were compared with each other in terms of lower inventory levels for an identical realised customer service level under rolling horizon planning. According to the results of the experimental study, the integrated model outperformed the sequential approach when the available capacity was strictly limited. If there were sufficient excess capacities, the integrated model, surprisingly, did not perform as well as the sequential approach, since it generated excess safety stocks by neglecting the re-planning opportunities found under rolling horizon planning. In this case, the re-planning opportunity adjusted integrated model was able to avoid the generation of excess safety stock and resulted in a more robust and promising performance than the sequential approach.

The comparison of the integrated model with the SDP revealed that, when we set a high target service level under an uncapacitated case, the absolute performance of the integrated model decreased since the integrated model yielded excess safety stocks. In the case of limited capacity, the lack of a full look-ahead capability of the integrated model to anticipate low and peak demand periods resulted in a higher gap to the theoretical lower bound.

The final problem addressed in this thesis was an S-GLSP. An optimal solution of this problem required a simultaneous determination of lot-sizes, detailed schedules and endogenous dynamic safety stocks. This problem had not yet been addressed in the literature. It was based on a real-world case study from a leading global company in the process industry. In this problem, various constraints and problem-specific assumptions, such as sequence-dependent setup times and costs, minimum production quantities and times, as well as inventory and production capacity constraints, were involved. Furthermore, the customer demand was serially correlated. We proposed several MILPs of different degrees of sophistication in order to address this complex problem.

First, within a deterministic APS optimisation model, a widely-used SA with the RoT specification of the exogenous safety stock targets was presented. Then, a new SA was introduced, where the exogenous safety stock targets were determined through a cost-optimisation method than an RoT. In this respect, two variants of a cDoS safety stock target for all products and a product-specific iDoS safety stock target were proposed. Afterwards, a new IM based on a cost-minimisation approach was developed. The proposed IM enabled the simultaneous determination of lot-sizes on a macro level, detailed schedules on a micro level as well as endogenous dynamic safety stocks during the non-integer, non-equidistant lengths of the replenishment cycles on a micro level. The IM was further adjusted to account for re-planning opportunities under rolling horizon planning. Finally, the IM-CRRL was presented, which extended the IM to take serially-correlated demand into account.

Based on a real-world dataset and a sensitivity analysis, the cost-saving potentials of the proposed approaches were evaluated under rolling horizon planning. The results showed that, within a deterministic APS optimisation model, replacing the widely-used RoT with a cost-optimised exogenous safety stock target provided substantial cost-saving potentials. Both the cost-optimised cDoS and iDoS approaches resulted in cost-savings of up to 10% over the RoT approach. Increasing the level of model sophistication by using the IM instead of the SA resulted in additional cost-savings of up to 20%. The results revealed that taking serially-correlated demand in the IM into account, i.e., the IM-CRRL, resulted in up to 5% cost-savings, despite its complexity.

To summarise, we briefly answer the research questions given in the introduction.

(i) We first introduced new MILPs to endogenously determine demand uncertainty parameters during the non-equidistant, discrete or continuous lengths of replenishment cycles for both the S-CLSP and the S-GLSP. Then, we applied bivariate linearisation techniques to endogenously set dynamic safety stocks by taking either a service level into account or applying the cost minimisation approach.

(ii) If capacity is strictly limited, using an integrated model instead of a sequential approach is beneficial. In the case of sufficient excess capacity, there is primarily enough flexibility to capture some of uncertainty by using re-planning opportunities under rolling horizon planning. In this case, an integrated model, which neglects re-planning opportunities, results in excess safety stocks. The integrated model can only outperform the sequential approach if it adjusts endogenous dynamic safety stocks according to re-planning opportunities.

(iii) We introduced a new MILP to take serially-correlated demand into account. Despite its complexity, this enabled additional cost-savings of up to 5% over the approach with the assumption of independent demand in our problem settings.

(iv) The industrial partner could achieve some quick cost savings of up to 10% within a deterministic APS optimisation model by simply switching from a widely-used rule-of-thumb approach for exogenous safety stock targets to a cost-based optimised approach, such as cDoS and iDoS. A further transition to the integrated model and the integrated model with serially correlated demand could mostly provide a much larger cost-saving potential of up to 20%.

6.2 Limitations and Future Research

This thesis addressed single-stage cases of stochastic lot-sizing and scheduling problems. In practice, it is also common to face multi-stage stochastic lot-sizing and scheduling problems. In a multi-stage case, the determination of the endogenous replenishment cycles does not solely depend on production periods, but also on the endogenous replenishment cycles during the other stages. Such interrelations complicate these types of problems significantly. If stochasticity is mainly introduced from the customer's side on final products it might in some cases be sufficient to place dynamic safety stocks only during the final stage. If this is true, then the integrated approaches proposed in this thesis can be easily extended to address such multi-stage problems. However, it is not yet clear how a multi-echelon dynamic safety stock placement problem can be fully integrated into the lot-sizing and scheduling problems. There is definitely an interest, both academically and practically, to investigate these types of integrated approaches in future research. The approaches proposed in this thesis can be a useful and promising staring point in this direction, too.

Using a shortage-cost minimisation approach in order to determine the dynamic safety stocks led to a trivariate non-linear function that depends on the inventory position and the mean and variance of the replenishment cycle demand. We were able to reduce this non-linear function to a bivariate one when demand was distributed normally. However, it is not yet clear how we can obtain a bivariate non-linear function when demand follows distributions other than the normal one. This would be another interesting topic to work on in the future, especially for in cases where demand is gamma distributed.

Another important aspect is the existence of re-planning opportunities under rolling horizon planning. In Chapter 4, we used a simple heuristic method to incorporate re-planning opportunities in terms of safety capacities in order to adjust endogenous dynamic safety stocks. In Chapter 5, we introduced a more comprehensive approach to capture re-planning opportunities under rolling horizon planning. Future research can focus on applying the latter approach in similar problems. It is a matter of interest to elaborately integrate the re-planning-opportunity adjustment coefficient into the integrated approaches without increasing the complexity. Nevertheless, in order to obtain optimal solutions, approaches like an SDP are required. Due to their computational limitations, however, these approaches are still less attractive in practice when it comes to addressing industrial problems that are usually large-scale and mainly problem-specific. Thus, improving the re-planning opportunity adjustment heuristic within an MILP system in future research would be more appealing to the practice.

Last but not least, the proposed ideas and approaches in this thesis addressing the S-CLSP and the S-GLSP can be applied in other types of the lot-sizing problems or in similar fields, such as the production planning stream, where lot-sizing or detailed scheduling is not the primary focus.

Bibliography

Aghezzaf, E. H., Jamali, M. A., Ait-Kadi, D., 2007. An integrated production and preventive maintenance planning model. European Journal of Operational Research 181 (2), 679–685.

Alaoui-Selsouli, M., Mohafid, A., Najid, N. M., 2012. Lagrangian relaxation based heuristic for an integrated production and maintenance planning problem. International Journal of Production Research 50 (13), 3630–3642.

Albey, E., Norouzi, A., Kempf, K. G., Uzsoy, R., 2015. Demand modeling with forecast evolution: An application to production planning. IEEE Transactions on Semiconductor Manufacturing 28 (3), 374–384.

Albey, E., Uzsoy, R., Kempf, K. G., 2016. A chance constraint based multi-item production planning model using simulation optimization. In: Proceedings of the 2016 Winter Simulation Conference. IEEE Press, pp. 2719–2730.

Allaoui, H., Castagliola, P., Pellerin, R., Sitompul, C., Aghezzaf, E. H., 2011. An integrated hierarchical production and maintenance-planning model. Journal of Quality in Maintenance Engineering 17 (3), 299–314.

Arts, J., Flapper, S., 2015. Aggregate overhaul and supply chain planning for rotables. Annals of Operations Research 224 (1), 77–100.

Askar, G., Sillekens, T., Suhl, L., Zimmermann, J., 2007. Flexibility planning in automotive plants. In: Günther, H. O., Mattfeld, D. C., Suhl, L. (Eds.), Management logistischer Netzwerke. Physica-Verlag Heidelberg, pp. 235–255.

Askar, G., Zimmermann, J., 2007. Optimal usage of flexibility instruments in automotive plants. In: Waldmann, K., Stocker, U. M. (Eds.), Operations Research Proceedings. Springer Berlin Heidelberg, pp. 479–484.

Askin, R. G., 1981. A procedure for production lot sizing with probabilistic dynamic demand. AIIE Transactions 13 (2), 132–137.

AT Kearney, 2016. 27th annual "State of Logistics Report". Council of Supply Chain Management Professionals. Oak Brook, IL, USA.

Bertrand, J., 2003. Supply chain design: Flexibility considerations. In: de Kok, T. G., Graves, S. C. (Eds.), Supply Chain Management: Design, Coordination and Operation. Vol. 11 of Handbooks in Operations Research and Management Science. Elsevier, pp. 133–198.

Bookbinder, J. H., H'ng, B. T., 1986. Rolling horizon production planning for probabilistic time-varying demands. International Journal of Production Research 24 (6), 1439–1458.

Bookbinder, J. H., Tan, J. Y., 1988. Strategies for the probabilistic lot-sizing problem with service-level constraints. Management Science 34 (9), 1096–1108.

Cachon, G., Terwiesch, C., 2012. Matching Supply with Demand: An Introduction to Operations Management, 3rd Edition. McGraw Hill.

Charnes, A., Cooper, W. W., 1959. Chance-constrained programming. Management Science 6 (1), 73–79.

Charnes, J. M., Marmorstein, H., Zinn, W., 1995. Safety stock determination with serially correlated demand in a periodic-review inventory system. Journal of the Operational Research Society, 1006–1013.

Copil, K., Wörbelauer, M., Meyr, H., Tempelmeier, H., 2017. Simultaneous lotsizing and scheduling problems: A classification and review of models. OR Spectrum 39 (1), 1–64.

Da Silva, G. C., Figueira, J., Lisboa, J., Barman, S., 2006. An interactive decision support system for an aggregate production planning model based on multiple criteria mixed integer linear programming. Omega 34 (2), 167–177.

de Kok, T. G., Fransoo, J. C., 2003. Planning supply chain operations: Definition and comparison of planning concepts. In: de Kok, T. G., Graves, S. C. (Eds.), Supply Chain Management: Design, Coordination and Operation. Vol. 11 of Handbooks in Operations Research and Management Science. Elsevier, pp. 597 – 675.

Disney, S. M., Gaalman, G. J., Hedenstierna, C. P. T., Hosoda, T., 2015. Fill rate in a periodic review order-up-to policy under auto-correlated normally distributed, possibly negative, demand. International Journal of Production Economics 170, 501–512.

Fleischmann, B., Meyr, H., 1997. The general lotsizing and scheduling problem. OR Spectrum 19 (1), 11–21.

Floudas, C. A., Lin, X., 2004. Continuous-time versus discrete-time approaches for scheduling of chemical processes: A review. Computers & Chemical Engineering 28 (11), 2109–2129.

Glasserman, P., 1997. Bounds and asymptotics for planning critical safety stocks. Operations Research 45 (2), 244–257.

Graves, S. C., 2011. Uncertainty and production planning. In: Kempf, K. G., Keskinocak, P., Uzsoy, R. (Eds.), Planning production and inventories in the extended enterprise. Vol. 151 of International Series in Operations Research and Management Science. Springer, pp. 83–101.

Groff, G. K., 1979. A lot sizing rule for time-phased component demand. Production and Inventory Management 20 (1), 47–53.

Helber, S., Sahling, F., Schimmelpfeng, K., 2013. Dynamic capacitated lot sizing with random demand and dynamic safety stocks. OR Spectrum 35 (1), 75–105.

Hemig, C., Rieck, J., Zimmermann, J., 2014. Integrated production and staff planning for heterogeneous, parallel assembly lines: An application in the automotive industry. International Journal of Production Research 52 (13), 3966–3985.

Jans, R., Degraeve, Z., 2007. Meta-heuristics for dynamic lot sizing: A review and comparison of solution approaches. European Journal of Operational Research 177 (3), 1855–1875.

Jans, R., Degraeve, Z., 2008. Modeling industrial lot sizing problems: A review. International Journal of Production Research 46 (6), 1619–1643.

Kang, Y., Albey, E., Hwang, S., Uzsoy, R., 2014. The impact of lot-sizing in multiple product environments with congestion. Journal of Manufacturing Systems 33 (3), 436–444.

Karimi, B., Ghomi, S. F., Wilson, J., 2003. The capacitated lot sizing problem: A review of models and algorithms. Omega 31 (5), 365–378.

Koçlar, A., 2005. The general lot sizing and scheduling problem with sequence dependent changeovers. Master's thesis, Middle East Technical University.

Koçlar, A., Süral, H., 2005. A note on "The general lot sizing and scheduling problem". OR Spectrum 27 (1), 145–146.

Lagodimos, A., Mihiotis, A., 2006. Overtime vs. regular shift planning decisions in packing shops. International Journal of Production Economics 101 (2), 246–258.

Lin, P. C., Uzsoy, R., 2016. Chance-constrained formulations in rolling horizon production planning: An experimental study. International Journal of Production Research 54 (13), 3927–3942.

Löhndorf, N., Minner, S., 2013. Simulation optimization for the stochastic economic lot scheduling problem. IIE Transactions 45 (7), 796–810.

Löhndorf, N., Riel, M., Minner, S., 2014. Simulation optimization for the stochastic economic lot scheduling problem with sequence-dependent setup times. International Journal of Production Economics 157, 170–176.

Manary, M. P., Willems, S. P., 2008. Setting safety-stock targets at Intel in the presence of forecast bias. Interfaces 38 (2), 112–122.

Mapes, J., 1993. The effect of capacity limitations on safety stock. International Journal of Operations & Production Management 13 (10), 26–33.

Meistering, M., Stadtler, H., 2017. Stabilized-cycle strategy for capacitated lot sizing with multiple products: Fill-rate constraints in rolling schedules. Production and Operations Management 26 (12), 2247–2265.

Merzifonluoğlu, Y., Geunes, J., Romeijn, H. E., 2007. Integrated capacity, demand, and production planning with subcontracting and overtime options. Naval Research Logistics 54 (4), 433–447.

Minner, S., 2000. Strategic Safety Stocks in Supply Chains. Lecture Notes in Economics and Mathematical Systems. Springer Berlin Heidelberg.

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

Najid, N. M., Alaoui-Selsouli, M., Mohafid, A., 2011. An integrated production and maintenance planning model with time windows and shortage cost. International Journal of Production Research 49 (8), 2265–2283.

Nam, S. J., Logendran, R., 1992. Aggregate production planning – A survey of models and methodologies. European Journal of Operational Research 61 (3), 255–272.

Neale, J. J., Willems, S. P., 2015. The failure of practical intuition: How forward-coverage inventory targets cause the landslide effect. Production and Operations Management 24 (4), 535–546.

Orcun, S., Uzsoy, R., Kempf, K. G., 2009. An integrated production planning model with load-dependent lead-times and safety stocks. Computers & Chemical Engineering 33 (12), 2159–2163.

Othman, M., Bhuiyan, N., Gouw, G. J., 2012. Integrating workers' differences into workforce planning. Computers & Industrial Engineering 63 (4), 1096–1106.

Powell, W. B., 2014. Clearing the jungle of stochastic optimization. Informs TutORials in Operations Research.

Ramezanian, R., Rahmani, D., Barzinpour, F., 2012. An aggregate production planning model for two phase production systems: Solving with genetic algorithm and tabu search. Expert Systems with Applications 39 (1), 1256–1263.

Ravindran, A., Kempf, K. G., Uzsoy, R., 2011. Production planning with load dependent lead times and safety stocks for a single product. International Journal of Planning and Scheduling 1 (1), 58–89.

Ray, W., 1980. The significance of correlated demands and variable lead times for stock control policies. Journal of the Operational Research Society 31 (2), 187–190.

Rebennack, S., Kallrath, J., 2015. Continuous piecewise linear delta-approximations for bivariate and multivariate functions. Journal of Optimization Theory and Applications 167 (1), 102–117.

Rossi, R., Kilic, O. A., Tarim, S. A., 2015. Piecewise linear approximations for the static–dynamic uncertainty strategy in stochastic lot-sizing. Omega 50, 126–140.

Schneeweiss, C., 2003. Distributed Decision Making. Springer Science & Business Media.

Sillekens, T., 2008. Aggregierte Produktionsplanung in der Automobilindustrie unter besonderer Berücksichtigung von Personalflexibilität. Ph.D. thesis, Universität Paderborn.

Sillekens, T., Koberstein, A., Suhl, L., 2011. Aggregate production planning in the automotive industry with special consideration of workforce flexibility. International Journal of Production Research 49 (17), 5055–5078.

Silva, J. P., Lisboa, J., Huang, P., 2000. A labour-constrained model for aggregate production planning. International Journal of Production Research 38 (9), 2143–2152.

Silver, E., 1978. Inventory control under a probabilistic time-varying, demand pattern. AIIE Transactions 10 (4), 371–379.

Silver, E. A., Bischak, D. P., 2011. The exact fill rate in a periodic review base stock system under normally distributed demand. Omega 39 (3), 346–349.

Silver, E. A., Pyke, D. F., Thomas, D. J., 2017. Inventory and Production Management in Supply Chains, 4th Edition. CRC Press.

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

Tarim, S. A., Kingsman, B. G., 2004. The stochastic dynamic production/inventory lot-sizing problem with service-level constraints. International Journal of Production Economics 88 (1), 105–119.

Tavaghof-Gigloo, D., Klosterhalfen, S. T., Minner, S., 2019. Integrated lot-sizing, scheduling and dynamic safety stock planning: A real-world case study from process industry. Working Paper.

Tavaghof-Gigloo, D., Minner, S., 2019. Planning approaches for stochastic capacitated lot-sizing with service level constraints. Working Paper.

Tavaghof-Gigloo, D., Minner, S., Silbermayr, L., 2016. Mixed integer linear programming formulation for flexibility instruments in capacity planning problems. Computers & Industrial Engineering 97, 101–110.

Tempelmeier, H., 2007. On the stochastic uncapacitated dynamic single-item lotsizing problem with service level constraints. European Journal of Operational Research 181 (1), 184–194.

Tempelmeier, H., 2011. A column generation heuristic for dynamic capacitated lot sizing with random demand under a fill rate constraint. Omega 39 (6), 627–633.

Tempelmeier, H., Herpers, S., 2010. ABC_β – A heuristic for dynamic capacitated lot sizing with random demand under a fill rate constraint. International Journal of Production Research 48 (17), 5181–5193.

Tempelmeier, H., Herpers, S., 2011. Dynamic uncapacitated lot sizing with random demand under a fillrate constraint. European Journal of Operational Research 212 (3), 497–507.

Tempelmeier, H., Hilger, T., 2015. Linear programming models for a stochastic dynamic capacitated lot sizing problem. Computers & Operations Research 59, 119–125.

Transchel, S., Minner, S., Kallrath, J., Löhndorf, N., Eberhard, U., 2011. A hybrid general lot-sizing and scheduling formulation for a production process with a two-stage product structure. International Journal of Production Research 49 (9), 2463–2480.

Urban, T., 2000. Reorder level determination with serially-correlated demand. Journal of the Operational Research Society 51 (6), 762–768.

Vielma, J. P., Ahmed, S., Nemhauser, G., 2010. Mixed-integer models for nonseparable piecewise-linear optimization: Unifying framework and extensions. Operations Research 58 (2), 303–315.

Walter, M., Sommer-Dittrich, T., Zimmermann, J., 2011. Evaluating volume flexibility instruments by design-of-experiments methods. International Journal of Production Research 49 (6), 1731–1752.

Weinstein, L., Chung, C., 1999. Integrating maintenance and production decisions in a hierarchical production planning environment. Computers & Operations Research 26 (10), 1059–1074.

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