Inventory Allocation Under Asymmetric Information and Decision Power

Inaugural Dissertation
to Obtain the Academic Degree of
a Doctor in Business Administration
at the University of Mannheim

submitted by

Christoph Löffel
Mannheim

Dean: *Joachim Lutz*
Referent: *Prof. Dr. Moritz Fleischmann*
Correferent: *Prof. Dr. Raik Stolletz*

Day of oral examination: February 6, 2026

To Sabine, Frank & Johannes.
And to Christina.

Acknowledgements

This dissertation marks the culmination of my twelve-year journey at the University of Mannheim. Studying in Mannheim has been an incredibly enriching experience, allowing me to grow not only intellectually but also personally. It is here that I discovered the field to which I wish to dedicate my career, and where I formed friendships that have supported and inspired me throughout the years.

First and foremost, I would like to express my deepest gratitude to Moritz Fleischmann. Moritz has not only been my PhD advisor but, together with Raik Stolletz, also introduced me to the fields of supply chain and operations management through their entry-level courses. Starting with a catchy "That's logistics" video shown at the end of the first class in fall 2014, Moritz's teaching and research approach helped me appreciate that supply chain management is truly everywhere around us and central to many societal challenges. His approach of combining rigorous analytical methods with practical relevance has always inspired me. It is this approach that has shaped my own way of working and ultimately motivated me to pursue a doctoral degree. I am immensely grateful that Moritz welcomed me as an external doctoral student at his chair. The skills and insights I have gained under his guidance will continue to benefit me for the rest of my life.

I would also like to thank Steffen Klosterhalfen. For me, he exemplifies how impactful academic work can emerge from an industry context. Ever since meeting him soon after joining the industry, I have admired his ability to drive projects that result in both academic publications and real-world implementations. His approach has deeply influenced my decision to pursue a doctoral degree alongside my industry work. I am especially thankful for his active collaboration on two of my projects and for the many insightful discussions we have shared.

My sincere thanks also go to Tobias Hausen, who not only hired me into a fantastic team early in my career but has also been an important mentor ever since. The discussions with him about the dynamics shaping supply chains in practice have been a major source of inspiration for the research we have conducted together over the past years.

I would also like to thank Pardis Sahraei for being a wonderful research partner on our joint project. I greatly valued the fresh perspectives she brought to our work, and her insights have meaningfully enriched my research.

My sincere thanks go to the examination committee for their time and effort in critically reviewing my work and for granting me the opportunity to defend my dissertation.

Finally, it is hard to put into words the depth of my gratitude to my family. Throughout my life, their unconditional support has made every new challenge feel surmountable. They have taught me the value of ambition and the courage to pursue new paths, while also reminding me of the importance of balance. Without their encouragement and belief in me, this journey would not have been possible. Lastly, I want to thank my partner Christina. Without her, this thesis might not exist. Throughout the last years, she has always kept me on track, provided me with encouragement, and offered the right perspective whenever needed.

Summary

Appropriately allocating inventories across the supply network is essential to a firm's success. To manage inventory effectively, a firm must navigate complex trade-offs and account for constraints such as limited production capacity, raw material availability, storage space, transportation capacity, and capital restrictions, all within an often uncertain business environment. Moreover, distributed decision-making and conflicting objectives among the involved actors further complicate inventory allocation in many real-world environments (Romano 2003, Tuomikangas and Kaipia 2014).

This thesis is based on a collaboration with a multinational agrochemical supplier and addresses three key inventory allocation challenges faced by the case company. Chapter II proposes an optimization-based budget planning model to support the annual budget planning cycle of the case company's business units, where the sales and supply chain departments align on a set of volume guarantees to be supplied along the planning horizon while ensuring adherence to a strict inventory investment cap imposed by the business unit head. While Chapter II considers a single business unit under a given maximum inventory investment ceiling, Chapter III focuses on how the company's headquarters should disaggregate a total inventory investment ceiling into business unit-specific ceilings under private information and conflicting incentives. To this end, we propose an auction mechanism through which business units can acquire a share of the total inventory investment budget. Chapter IV addresses the question of how much of which final product the headquarters of the firm should produce from a limited set of raw materials. The allocation problem is complicated by uncertain customer demand and private information held by local product managers who optimize local instead of global profits. We propose a two-step allocation mechanism that combines an initial top-down allocation with subsequent bottom-up adjustments facilitated through transfer payments.

Table of Contents

List of Figures

List of Tables

Chapter I

Introduction

Allocating the right amount of inventory to each point in a firm's supply network is a critical challenge for corporations (Cachon and Lariviere 1999b). Effective inventory management is essential for navigating the complex trade-offs between minimizing costs from manufacturing, procurement, and warehousing, reducing total capital employed, and mitigating the risk of lost sales due to stockouts. When addressing these trade-offs, firms must consider various constraints such as limited production capacity, raw material availability, storage space, transportation capacity, and capital restrictions. Determining optimal inventory allocation decisions in real-world environments is often further complicated by conflicting goals and distributed decision-making (Mishra et al. 2007).

In modern enterprises, corporate success is thereby no longer evaluated solely based on profitability. Instead, stakeholders rely on a diverse set of financial and non-financial metrics to assess firm performance. Consequently, management teams must align operational decisions with a multidimensional scorecard of potentially conflicting performance metrics (Bhagwat and Sharma 2007).

This dissertation is based on an industry collaboration with a multinational agrochemical supplier. At the case company, as in many real-world settings, decision-making power is distributed across different departments. Effective collaboration among functions such as sales, marketing, production, finance, and supply chain management is critical, but often challenged by asymmetric information and conflicting incentives (Romano 2003, Tuomikangas and Kaipia 2014). The case company operates a multi-echelon, long-lead-time supply chain to meet the global demand for crop protection products. Demand in this industry is highly seasonal and uncertain, influenced by hard-to-predict factors such as local disease pressure and crop prices (Hausen et al. 2006). Long supply lead times, limited production capacities, and high demand uncertainty contribute to high inventory levels and capital employment, making effective inventory management a critical lever for maximizing expected profits.

In addition, the case company considers inventory not only as part of its cost function, but as a key performance indicator (KPI) in its own right (Agrawal and Osadchiy 2024, Lai and Xiao 2018). Financial analysts commonly use the total cash invested in inventories as an independent measure to evaluate the firm's operating efficiency and benchmark it against core competitors (Hendricks and Singhal 2009). As a result, the management team considers inventory investment levels as an independent KPI that must be actively managed to meet investor expectations and those of the broader financial market.

The company's global operations are divided into distinct functional areas, each characterized by unique decision-making power, proprietary information, and incentive structures. This dissertation studies three specific inventory allocation problems faced by the case company. For each of them, we identify the actors involved, their roles and incentives, and the coordination mechanisms currently in place. We then propose solution approaches that are implementable in the decision-making context of the case company, evaluate their performance against benchmarks based on synthetic and real-world data, and study the resulting solution structure.

Chapter II (joint work with Steffen Klosterhalfen, Tobias Hausen, and Moritz Fleischmann) addresses a critical challenge faced by the case company: aligning the commercial plans of the sales department with the supply plans of the supply chain department. At the case company, cross-functional coordination is achieved through annual Sales and Operations Planning (S&OP) budget meetings. During this process, the sales and supply chain organizations agree on a set of supply volume guarantees to align commercial and production plans. Our research focuses on determining these volume guarantees. To limit total inventory investment while minimizing expected lost margins, the business unit head imposes a maximum allowable inventory investment that the organization must not exceed.

We develop an optimization-based budget planning model to support the S&OP negotiation process. The model explicitly accounts for differences in decision-making scope and authority among relevant actors. Furthermore, it incorporates flexibility for future plan adjustments through an affinely adjustable robust optimization framework.

Based on two numerical studies, we find that optimizing volume guarantees can substantially reduce expected lost margins compared to uniform guarantees. For our real-world data set, our approach saves up to 32% relative to a current practice benchmark.

While the first essay examines how to coordinate the sales and supply chain departments under an exogenously given total inventory investment ceiling, the second essay (joint work with Steffen Klosterhalfen and Moritz Fleischmann), presented in Chapter III, addresses the higher-level planning problem of determining the business-unit-specific inventory investment ceilings. Specifically, we consider a corporation composed of multiple independently managed business units overseen by an overarching management board. Prior to the start of the business year, the management board sets a total inventory investment ceiling for the corporation and subsequently disaggregates this ceiling into business unit-specific upper limits. However, private information held by self-optimizing business unit heads complicates this allocation task. As business units compete for a higher share of the total inventory investment ceiling, they have an incentive to strategically withhold or distort private information. Our work focuses on how to best distribute a given total inventory investment ceiling across business units.

The case company currently applies an allocation heuristic that relies solely on information centrally available to the management board. While this heuristic avoids adverse selection between the management board and its business units, its performance implications have been questioned by both parties.

As an alternative, we propose an auction-based allocation mechanism that transforms the inventory investment ceiling into tradable inventory permits, which are then auctioned among business unit heads. We analyze the optimal decision-making of business unit heads and provide closed-form solutions. The proposed mechanism is evaluated through two numerical studies based on synthetic data and real-world data from the case company. Results indicate that the auction-based approach improves expected total profit on average by 4% compared to the firm's current allocation method. Furthermore, we uncover non-monotonic relationships between business unit characteristics, inventory ceilings, and supply quantities, highlighting the importance of leveraging private information through market-based mechanisms rather than relying solely on limited central data when determining business unit-specific inventory investment ceilings.

In Chapters II and III, we examine how to allocate a limited inventory investment budget across products and business units. In Chapter IV (joint work with Pardis Sahraei and Moritz Fleischmann), we turn to another critical allocation problem faced by the case company, namely which final products to produce from a limited volume of active ingredients to maximize the expected profits of the firm.

The production of crop protection products follows a two-stage process which is centrally managed by the headquarters. The first stage involves synthesizing active ingredients in large, capital-intensive facilities, where high utilization rates are essential for cost efficiency. To prepare for the highly seasonal demand for crop protection products, the company must ramp up active ingredient production approximately a year in advance (Shah 2004). Since expanding capacity requires long-term investments (Fritz and Hausen 2009), available active ingredient volumes remain fixed in the short and medium term. Shortly before the selling season, headquarters initiates the second production stage where one or more active ingredients are combined into finished products. At this point, headquarters faces a complex allocation decision: selecting which final products to manufacture given limited active ingredient volumes to maximize expected profits. To decide on how much of which final product to produce, the case company relies on orders from local product managers who are the only ones with direct market access. This information asymmetry, combined with an incentive structure that encourages product managers to prioritize their own local over global profits, complicates the allocation problem. Currently, the company applies a lexicographic allocation approach that prioritizes orders from product managers based on their profitability. While this approach discourages strategic behaviors such as order inflation, its appropriateness has been questioned. In particular, product managers who suffered significant lost sales have criticized the current approach after observing other, often only marginally more profitable, product managers carrying large unsold inventories into the next season.

To address these concerns, we propose a hybrid mechanism denoted as *base-and-refine*, which combines an initial top-down allocation with a refinement step that allows product managers to trade portions of their allocations through transfer payments based on centrally set transfer prices. Our numerical experiments demonstrate that our proposed method consistently outperforms both theoretical and company benchmarks, significantly narrowing the gap to the first-best solution achievable under full information. Moreover, we show that the two-step design, integrating top-down allocation with transfer-price-based adjustments, delivers superior performance compared to either step alone and is more robust to transfer price misspecification than a pure transfer-pricing approach.

Chapter II

Decision support for sales and operations planning under asymmetric power: A case study from the agrochemical industry

with Steffen Klosterhalfen, Tobias Hausen, and Moritz Fleischmann

Abstract

Agrochemical companies operate multi-echelon, long-lead-time supply chains to serve seasonal and uncertain demand for crop protection products from farmers around the globe. To match supply and demand in this challenging setting, alignment of the sales and supply chain functions is crucial. We investigate this alignment task, building on a case study. At the case company, cross-functional coordination is achieved through annual sales and operations planning budget meetings. In the budgeting process, sales and supply chain organizations agree on a set of supply volume guarantees to align commercial and production plans. The guarantees must respect a maximum allowable inventory investment imposed by the business unit head. Our work focuses on the choice of these volume guarantees. Specifically, we support the S&OP negotiation process by developing an optimization-based budget planning model. Importantly, the model reflects differences in the decision-making scope and power of the relevant actors. Moreover,

it captures the available flexibility of future plan adjustments by means of an affinely adjustable robust optimization approach. We evaluate the performance of our model against relevant benchmarks in two numerical studies, based on synthetic and real-world data. We also study the structure of the obtained solution. We find that optimizing the volume guarantees can substantially reduce lost margins, relative to uniform guarantees. For our real-world data set, our approach saves up to 32% relative to a current practice benchmark.

2.1. Introduction

A company's success depends on the efficient and effective collaboration of its individual functions, such as sales, marketing, production, finance and supply chain. However, cross-functional alignment is often hampered by opposing incentives, asymmetric information and distributed decision-making power (Romano 2003, Tuomikangas and Kaipia 2014). Sales and operations planning (S&OP) meetings are a means of coordination that many companies have put in place to align their mid-term commercial aspirations with given operational constraints (Thomé et al. 2012, Oliva and Watson 2011). Although a substantial amount of literature has been published on various aspects of S&OP, such as its procedural setup and empirical effectiveness (Tuomikangas and Kaipia 2014), research on analytical models that support S&OP decision-making is limited (Pereira et al. 2020). Moreover, many analytical contributions assume, at least implicitly, a single, central decision-maker, optimizing a single objective function. By doing so, they ignore the very reason for establishing an S&OP process, namely, conflicting incentives as well as differences in information availability and decision-making power. Consequently, such model formulations face two major challenges that limit their value in practice. First, asymmetric and incomplete information hinders appropriate model parameterization. Second, any globally optimized solution that conflicts with the incentives of strong decision-makers will likely be blocked and thus not be implementable.

In this paper, we consider a real-world S&OP problem from the agrochemical industry that is characterized by the above-mentioned incentive, information, and decision-making power challenges. The case company operates a global, multi-echelon, capacitated, long-lead-time supply chain to serve the seasonal and uncertain demand for agrochemical products. To align the commercial plan of the sales organization with the production and distribution plan of the supply chain, both functions agree on a set of supply volume guarantees. The guarantees must respect a maximum allowable inventory investment constraint imposed by the business unit head to steer working capital and cash flow requirements. At present, the volume guarantees at the case company result from an unstructured and iterative negotiation process, which employs rather simple allocation strategies. We propose an alternative optimization-based approach to determine these guarantees. By closely reflecting the existing power and incentive structure at the case company, we ensure that the identified solution is implementable in practice.

Our paper is rooted in the S&OP literature. Our main contributions are fourfold. First, we provide insights into a real-life S&OP problem in the agrochemical industry. We discuss the actors involved, their roles, incentives, and decision power, as well as the coordination mechanism that is currently in place at the case company. We point out that the setting is at odds with common assumptions in the literature concerning differences in power, information availability, and objectives between key stakeholders. We argue that these observations extend beyond our specific case and call for a critical reflection on how to model S&OP decisions.

Second, we disentangle the elements of a currently unstructured negotiation process governing the interplay between sales, production, and distribution. We derive a conceptual model of this process, thereby enhancing transparency and providing a basis for a formal mathematical model.

Third, we formulate an advanced optimization model that captures core S&OP elements in their organizational context. We propose to reflect the distribution of decision-making power within the case company through a nested optimization model. The inner problem captures the task of the supply chain department to supply volumes to the sales department while guaranteeing robust adherence to an exogenous maximum inventory investment target introduced by the business unit head under uncertain market demands. We argue that adjustable robust optimization is a well-suited methodology for this task. The outer problem represents the decision of the sales department. The task of the sales department is to pick a volume plan which minimizes expected lost margin and is suppliable under the maximum inventory investment constraint managed by the supply chain department. To accurately reflect the different decision scopes of both departments, we therefore nest a robust minimization problem (supply chain department) as a constraint in an expected value minimization problem (sales department). This distinguishes our work methodologically from the existing S&OP literature. While the details of S&OP differ between companies, we believe that the arguments for our methodological choice apply more broadly and may help to add realism to the modeling of S&OP in general.

Fourth, we test our model numerically, and compare it against relevant benchmarks relying on simpler heuristics. Using a controlled synthetic data set, we identify key performance drivers. We then calibrate the observed effects on the basis of a real-world data set from the case company.

The remainder of this paper is structured as follows. In the next section, we provide an overview of the related literature. In Section 2.3, we present relevant aspects of agrochemical supply chains, as well as the budget planning problem within the case company's S&OP process. We develop our analytical model formulation in Section 2.4. In Sections 2.5 and 2.6, we assess the performance of our approach based on synthetic and real-world data from the case company, respectively. We conclude our work by discussing limitations and opportunities for future research in Section 2.7.

2.2. Literature review

The concept of S&OP has received substantial attention since its emergence in the 1980s (Wallace and Stahl 1999). Although S&OP cannot dispel cross-functional conflicts, its goals are to establish a temporal consensus on mid-term priorities based on overarching strategic plans and to obtain a single and aligned tactical plan, balancing the sales aspirations of the commercial functions with the supply and distribution capabilities of operations (Goh and Eldridge 2019). We contribute to the S&OP literature by providing practical insights into processes and coordination mechanisms used to facilitate S&OP at a multinational agrochemical company.

As opposed to the rich research streams covering the procedural setup and the empirical analyses of S&OP (Grimson and Pyke 2007, Gunasekaran et al. 2001, Kristensen and Jonsson 2018), the literature on analytical models to guide S&OP decisions is sparse. Moreover, these works mostly focus on extending traditional production planning problems with selective decisions of a single additional function. We refer the reader to Pereira et al. (2020) for a comprehensive overview. Only a few authors have proposed holistic model formulations that fully integrate the supply and sales decisions commonly addressed by S&OP. In the majority of the existing contributions, optimal solutions are obtained by consolidating the performance trade-offs of different cross-functional decisions into a single, integrated objective function of a central decision-maker.

In this vein, Nemati and Alavidoost (2019) provide a fuzzy mixed-integer program (MIP) to jointly optimize purchasing, production, distribution, and sales planning at an Iranian fast-moving consumer goods company. Susarla and Karimi (2018) combine procurement, multi-echelon production and distribution planning, and cross-market sales. They develop a deterministic MIP formulation to optimize the total profits of a multinational pharmaceutical supply chain. Based on a case study from an oriented strand board company in Canada, Feng et al. (2008) present an MIP capturing decisions of sales, production, distribution, and procurement in a deterministic make-to-order environment, which they later extend to the stochastic case (Feng et al. 2013). Bajgiran et al. (2016) provide a deterministic MIP to support the S&OP problem of a Canadian lumber supply chain that demands joint optimization of harvesting, procurement, production, inventory, transportation, and sales activities. Motivated by the planning problem of an automotive company, Torabi and Hassini (2009) capture procurement, production, distribution, and sales decisions in a single, integrated fuzzy goal programming formulation. For a comprehensive value-based view on S&OP models, Hahn and Kuhn (2011) combine the physical product flow driven by overtime, production, transportation, and marketing campaign plans with the financial perspective of associated cash flows, financing costs, and constraints in a deterministic linear program (LP).

Assuming that all cross-functional decisions can be jointly optimized through a single global objective function does not capture the reality of distributed decision-making power and conflicting cross-functional incentives. These are the core arguments for the importance and establishment of S&OP, however (Oliva and Watson 2011). Very few authors explicitly address distributed, asymmetric decision-making in their model formulations (Rius-Sorolla et al. 2020). Hu et al. (2011) model conflicting incentives and the distributed decision-making power of production and sales departments through Nash and Stackelberg games and assess its implications on the outcomes of a generic tactical production and distribution planning problem. Based on a case study from the semiconductor industry, Karabuk and Wu (2003) propose a capacity planning model under distributed decision-making that incorporates information asymmetry and conflicting incentives of manufacturing, marketing, and the divisional head in a stochastic program (SP). To capture the impact of distributed tactical decision-making under conflicting objectives, Hjaila et al. (2017) integrate game-theoretical concepts in a mathematical programming formulation to determine production, distribution, and sales plans for an energy production and chemical manufacturing supply chain. Our paper contributes to the latter research stream on S&OP models that explicitly reflect distributed decision-making and commonly perceived organizational behavior in their model formulations.

Scholars have studied various supply chain planning problems in the agrochemical industry, both in the crop protection and in the seeds business. Bassett and Gardner (2010) discuss the network design, as well as monthly production and shipping problem of an agrochemical supply chain. The authors propose an integrated mixed integer linear program to optimize total profitability and apply it to a single product of Dow AgroSciences. Naraghi and Jiang (2023) address a production and distribution planning problem of an agrochemical active ingredient under uncertain demand. The authors develop a stochastic mixed-integer nonlinear programming model to minimize total supply costs and expected lost demand. Motivated by the inventory management of crop protection products characterized by highly seasonal and uncertain demand, Schlapp et al. (2022) examine how firms should determine both inventory quantity and timing to maximize expected profits. Their work introduces a time dimension to the classical newsvendor problem. Bansal and Nagarajan (2017) address the production planning problem of a seeds supplier facing limited raw material availability and uncertain production yields. The authors analyze the option of investing in an additional production location that serves as a backup in the case of low production yield. Also in the seeds business, Bansal and Dyer (2020) develop an optimization approach for an inventory planning problem for a product portfolio in a newsvendor-like environment with demand substitution. Comhaire and Papier (2015) study the seed production network of Syngenta subject to supply- and demand-side uncertainties. The authors propose a simulation tool to improve the expected profit of the firm. Our study contributes to the literature on agrochemical supply chain planning by providing a real-world case study of the S&OP process of a multinational agrochemical supplier.

Based on the research collaboration, we propose a decision-support model for S&OP in the agrochemical industry. At the case company, supply and sales plans must be aligned with an overarching maximum allowable inventory investment through a shared budget plan that consists of a set of volume guarantees to be supplied by the supply chain organization to the sales regions. The agreed volume guarantees must comply with the maximum allowable inventory investment even under uncertain demand. We ensure this requirement by choosing a robust optimization formulation for modeling the multi-stage production of agrochemical products. We refer to Ben-Tal et al. (2009), Bertsimas et al. (2011) and Gorissen et al. (2015) for an overview of robust optimization. This modeling approach distinguishes our work methodologically from that in the existing literature. To appropriately capture the existing flexibility to adjust decisions over time in response to revealed uncertainties and thus to avoid overly conservative assessments of worst-case performance through a traditional static model, we use an affinely adjustable robust optimization (AARO) formulation based on Ben-Tal et al. (2004). Over the last two decades, AARO approaches have been applied to various operations research problems, such as robust inventory management in multi-stage supply chains (Aharon et al. 2009, Kim and Chung 2017), strategic network planning (Ning and You 2017) and cross-organizational coordination (Ben-Tal et al. 2005). For an overview on adjustable robust optimization, see Yanıkoğlu et al. (2019).

2.3. Problem description

Farmers use crop protection products (CPPs) to safeguard production yields by preventing crop damage through weeds (herbicides), insects (insecticides) or fungi (fungicides) on their fields (Agovino et al. 2019, Oerke and Dehne 2004). As both the research and production processes of agrochemicals are capital intensive, the global market of CPPs is served by a small number of multinational organizations operating global supply networks.

Similar to pharmaceuticals (Shah 2004), CPPs rely on active ingredients (AIs) to provide the desired crop protection effects. AIs are synthesized in large-scale reactors and typically produced at global production sites. While this exploits economies of scale and makes complex chemical production economically viable, it also renders the supply chain less responsive. Capacity and therefore AI volumes are inflexible. Moreover, the lead times to the next production step, referred to as formulation, are long. In the formulation stage, AIs are blended with several formulants (inert ingredients) and processed into the final bulk product, which is then filled, packaged, and labeled in compliance with local regulations into a marketable sales article (Gehen et al. 2019). In contrast to AI production, the formulation and subsequent tasks are conducted in regionally distributed sites from where the final products are sold through one or multiple distribution layers to farmers (Fritz and Hausen 2009).

The demand for CPPs is both highly seasonal and uncertain. To successfully meet this demand with the above-described long lead time and capacitated supply chain, thorough cross-functional coordination is crucial. This is commonly achieved through S&OP. At the case company, S&OP processes are composed of a cascade of meetings held regularly throughout the year at different hierarchical levels with various regional and product portfolio scopes.

Most prominently and to provide all lower-level S&OP activities with an overarching directive, an annual budget meeting is hosted by the business unit (BU) head, together with the senior management of the relevant functions, namely, sales and supply chain. As an integral part of the process, a budget plan is developed to coordinate the commercial ambitions of the sales department (SD) with the supply capability of the supply chain department (SCD) for the upcoming seasons. The alignment is complicated by several distinct functional business challenges, both on the demand and the supply side.

On the demand side, the application of CPPs by farmers is limited to only a few weeks per year. It is affected by various hard-to-predict factors, such as local weather conditions driving the disease pressure on the fields, fluctuating prices for agricultural commodities, and the sales activities of competitors. Due to this high demand seasonality and uncertainty, the SD is enticed to adjust sales plans dynamically in response to current but volatile regional market conditions.

On the supply side, the constrained production capacity as well as the long production and distribution lead times demand pre-production and inventory build-up prior to the selling season. This limits the SCD's ability to react and adjust to changing in-season sales plans (Affonso et al. 2008, Lahloua et al. 2018).

Acknowledging both perspectives, the SD and the SCD therefore agree during the budgeting process on a set of monthly volume guarantees to be supplied over the next two years. These guarantees serve as tactical guardrails within which sales teams can structure their marketing activities across the product portfolio, and the supply chain can detail production and distribution planning (Thomé et al. 2012, Romano 2003).

Prior to any budget discussions on how to allocate guarantees across the product portfolio, the BU head imposes a maximum allowable inventory investment target. This target serves two purposes. First, it can be used to actively control the "operating mode" of the organization. The chosen target level impacts the trade-off between high inventory investment associated with high volume guarantees and the risk of lost margins. Second, it is a means to introduce aspects to the budgeting process that are neither directly SD-nor SCD-related. Such factors include the current competitive situation of the global crop protection market, strategic growth aspirations and/or potential working capital and cash flow requirements imposed by the board of directors.

To understand the role and incentives of the SD during the budgeting process, we highlight the implemented key performance indicators (KPIs) and the associated compensation schemes for sales managers. On the organizational level, the performance of the overall SD is predominantly measured based on the generated revenue and, most importantly, the associated contribution margin. To align the efforts of individual sales managers with the SD target, their compensation packages strongly depend on their individual contribution to those KPIs. Therefore, the SD has an incentive to capture as much of the seasonal and uncertain customer demand as possible. Consequently, it is in the SD's interest to strategically request volume guarantees that minimize the risk of lost margins, irrespective of other related costs caused by producing and stocking the requested quantities. The ability of noncommercial functions, such as the SCD, to challenge communicated sales aspirations is limited. The SCD does not have detailed information on local market conditions or on the overall competitive environment, nor does it have direct access to customers. This puts the SD in a stronger position to influence the outcome of the budgeting process.

Expectations toward the SCD are twofold. First, it needs to secure the supply of the volume guarantees agreed on during the budgeting process. Second, it needs to ensure that the maximum allowable inventory investment set by the BU head is not violated. One particular challenge is that the required inventory depends not only on the tactical supply plan but also on the available flexibility to adjust this plan over time in response to observed sales. Both requirements are reflected in the compensation scheme of the SCD at the case company. Supply chain managers receive a lump sum bonus payment at the end of the business year if the total inventory investment did not exceed the defined target and if the agreed volume guarantees were reliably supplied to the SD. Note that while the sales managers' compensation depends proportionally on generated contribution margin, the SCD receives a lump sum bonus for satisfying supply and inventory-related conditions. Our modeling approach presented in the next section reflects this difference in compensation between the SD and the SCD.

It is noteworthy that the BU head at the case company opted not to design a mechanism that aligns the incentives of the SD and the SCD to jointly minimize expected lost margin while adhering to the inventory investment target. Instead, the case company operationalized both performance targets separately by incentivizing the SD to focus on margin optimization and the SCD to prioritize inventory control. This strategy diverges from a mechanism design approach, which would seek to align agent incentives with the overall objectives of the firm (Groves and Loeb 1979). Discussions with the case company revealed that incorporating conflicting incentives in the S&OP process is viewed as desirable and is intended to foster accountability within each department's span of control. The S&OP process is thereby conceptualized as an arena in which interdepartmental conflicts can be constructively negotiated, particularly prior to and during the budget meeting. As the alignment process is iterative, unstructured and highly nontrivial, analytical support can help to guide the discussions and to limit the risk of subpar supply guarantees.

Based on the above characteristics, the goal of the budget planning process can be formulated as follows:
Given the set of potential volume guarantees that the supply chain organization can offer while respecting the inventory ceiling, select the guarantee set that minimizes the expected lost margin over the planning horizon, given uncertain sales.

Figure 2.1 visualizes this process with notation to be introduced in Section 2.4.

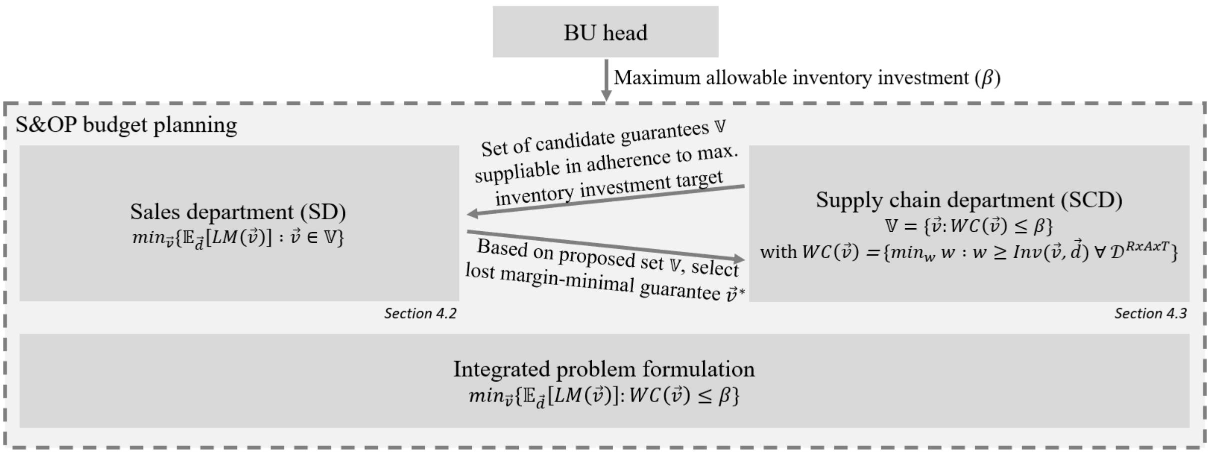

Figure 2.1.: Overview of proposed S&OP budget planning model subject to asymmetric decision-making power.

2.4. Model formulation

In this section, we propose a model of the outlined budget planning problem that considers the highlighted asymmetries in incentives and decision-making power between the SD and the SCD. The proposed model framework captures the power of the SD to choose a set of margin-oriented supply volumes as well as the SCD's task to assess feasibility with respect to both capacities and the admissible inventory investment ceiling. While capacity feasibility is a deterministic constraint, inventory levels result from uncertain demand, and adherence to the imposed ceiling has to be guaranteed for any realization. We model this guarantee as a robust, i.e. worst-case oriented, inventory investment minimization problem which the SCD has to solve to assess feasibility of a given set of supply volume guarantees. We then model the interaction between the SD and the SCD during the budgeting process by nesting the robust optimization subproblem faced by the SCD as a constraint in the expected lost margin minimization problem solved by the SD (see problem (2.3) to (2.6)). We formalize this modeling framework in Subsection 2.4.1. In Subsections 2.4.2 and 2.4.3, we then specify in detail how to model the expected value minimization problem of the SD and the worst-case problem faced by the SCD, respectively.

2.4.1. S&OP budget planning model

To coordinate mid-term sales and production plans, the SD and the SCD agree on a set of monthly AI volume guarantees to meet uncertain customer demand $d_{r,a,t}$ for AI $a \in \mathcal{A} = \{1, ..., A\}$ in sales region $r \in \mathcal{R} = \{1, ..., R\}$ in month $t \in \mathcal{T} = \{1, ..., T\}$. Let $v_{r,a,t}$ denote the volume that the SCD guarantees for AI a in region r and month t. A complete set of guarantees is denoted as

$$\vec{v} = (v_{1,1,1}, \dots, v_{R,A,T}). \tag{2.1}$$

A set of guarantees $\vec{v}$ is feasible if it respects available production capacities and the resulting total inventory investment $Inv(\vec{v}, \vec{d})$ does not exceed the imposed admissible level β for any possible demand realization $\vec{d} \in \mathcal{D}^{R \times A \times T}$. The SCD can determine the worst-case, i.e., maximum, total inventory investment resulting from the supply of $\vec{v}$, denoted by $WC(\vec{v})$, by solving the following robust optimization problem:

$$WC(\vec{v}) = \min_{w} w : w \geq Inv(\vec{v}, \vec{d}) \quad \forall \vec{d} \in \mathcal{D}^{R \times A \times T}. \tag{2.2}$$

Feasibility of $\vec{v}$ can then be expressed as $WC(\vec{v}) \leq \beta$.

As discussed in the previous section, the SD chooses the feasible volume guarantee set that minimizes the expected lost margins. Using the above notation, we can therefore formalize the considered S&OP budget planning problem as follows (comp. Fig. 2.1):

$$\min_{\vec{v}} \mathbb{E}_{\vec{d}}[LM(\vec{v})] \tag{2.3}$$

$$\text{s.t.} \quad WC(\vec{v}) \leq \beta \tag{2.4}$$

$$\text{with} \quad WC(\vec{v}) = \min_{w} w \tag{2.5}$$

$$\text{s.t.} \quad w \geq Inv(\vec{v}, \vec{d}) \quad \forall \vec{d} \in \mathcal{D}^{R \times A \times T}. \tag{2.6}$$

2.4.2. Lost margin estimation

To decide which of the feasible volume guarantees ($\vec{v}$) to choose, the SD must estimate the expected total lost margin for each of the candidate guarantees in objective function (2.3) of the budget planning problem, where we assume that the SD does not consider guarantees exceeding the upper support of the demand distribution. For any $\vec{v}$, the expected total lost margin across regions and AIs along the planning horizon is obtained as follows:

$$\mathbb{E}_{\vec{d}}[LM(\vec{v})] = \sum_{r=1}^{R} \sum_{a=1}^{A} \sum_{t=1}^{T} m_{r,a} \cdot \mathbb{E}_{\vec{d}}((d_{r,a,t} - v_{r,a,t})^{+}), \tag{2.7}$$

where $m_{r,a}$ denotes the per unit margin of met customer demand $d_{r,a}$. During the budget meeting, any monthly demand exceeding the agreed monthly AI volume guarantee is assumed to be lost. While during the execution phase of the budget plan, along the business year, the SCD and the SD might work together to try and identify and supply potential excess inventories to markets experiencing demand levels beyond the agreed guarantees, such interventions are addressed on shorter-term planning levels, below the discussed budget planning problem.

2.4.3. Worst-case inventory assessment

While the SD focuses on minimizing the expected total lost margin, the SCD is responsible for assessing whether a given set of volume guarantees is compatible with the imposed maximum allowable inventory investment. For this assessment, the SCD solves the subproblem (2.5) - (2.6) of the budget planning problem.

Any given volume guarantee can be realized via various supply plans with different associated inventory investments. In Subsection 2.4.3, we first present a deterministic supply chain model to minimize the total inventory investment linked to the supply of a given set of volume guarantees $\vec{v}$ under certain demands.

In reality, customer demand is, however, uncertain. To ensure that the volume guarantees can be supplied in adherence to the imposed inventory budget under any realization of market demand, the SCD must identify a tactical production and distribution plan that minimizes the associated worst-case total inventory investment. We model this planning task as an RO problem in Subsection 2.4.3 and extend it to an AARO formulation in Subsection 2.4.3 to account for existing SC flexibilities. For an overview of techniques available for formulating and solving robust models, we refer to Ben-Tal et al. (2009).

Deterministic supply chain model

We model the global AI production and the subsequent regional distribution over the next T months as follows. First, A active ingredients are synthesized in large central and dedicated sites able to continuously produce AI a in month t ($P_{a,t}$) at a monthly maximum capacity of c_a. Produced AIs can be stored at the production sites. We denote the corresponding inventory level of AI a at the end of month t by $I^{glo}_{a,t}$. This inventory is valued at a unit cost of goods of p^{glo}_a.[1] To ensure sufficient inventory coverage beyond period T, at least $i^{glo}_{a,T}$ units of AI a must be held in global storage at the end of the planning horizon. Second, regions $r \in \mathcal{R}$ order AIs from central storage facilities. We refer to the order of region r for AI a at the beginning of month t as $O_{r,a,t}$. After order placement, AIs are shipped from the central storage facilities to the regions, with a distribution lead time of $l_{r,a}$ months.

[1] Note that although the case of multiple production facilities per AI could be covered by the model through the introduction of an additional decision variable index, we directly present the single AI facility model that applies to the setting of the case company.

To ensure sufficient time for regional formulation, filling, packaging, and distribution, the SCD steers regional AI availability through minimum stock level requirements. Therefore, the regional inventory of AI a in region r at the end of month t, $I^{reg}_{r,a,t}$, must cover at least the AI volume guarantees of the next $\lambda_{r,a}$ months at all times. In addition, at least $i^{reg}_{r,a,T}$ units of AI a must be held in stock in region r at the end of the planning horizon. Similar to those at central storage facilities, regional stocks and in-transit inventories are valued at a unit cost of goods of $p^{reg}_{r,a}$.

Finally, the SD fulfills regional market demands up to the monthly volume guarantee aligned with the SCD during budgeting. We summarize the presented supply chain setup in Figure 2.2.

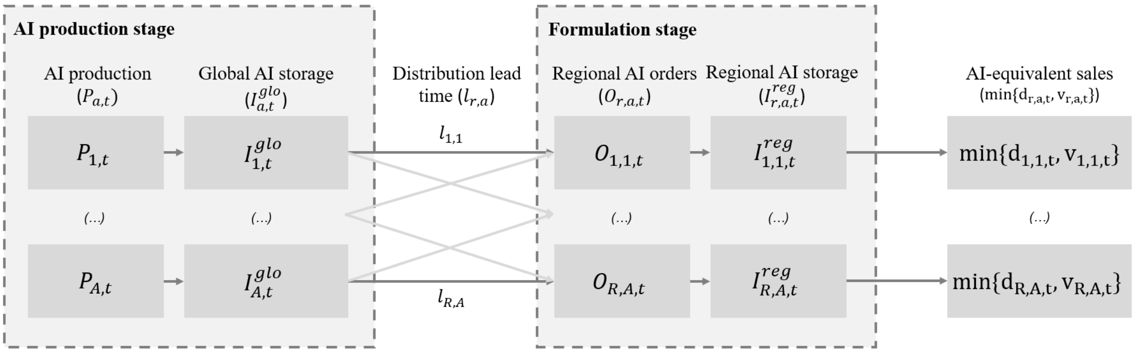

Figure 2.2.: Schematic overview of the modeled agrochemical supply chain.

Based on the described supply chain setup, the following deterministic linear program can be formulated to determine the minimal total inventory investment ($Inv(\vec{v}, \vec{d})$) for a given set of volume guarantees under deterministic market demands:

$$\min\ Inv(\vec{v},\vec{d}) = \min\ F = \sum_{a\in\mathcal{A}}\sum_{t\in\mathcal{T}}(p_a^{glo}I_{a,t}^{glo} + \sum_{r\in\mathcal{R}} p_{r,a}^{reg}(I_{r,a,t}^{reg} + \sum_{i=0|t+i\leq T}^{l_{r,a}-1} O_{r,a,t})) \tag{2.8}$$

$$\text{s.t. } P_{a,t} \leq c_a \qquad \forall\, a\in\mathcal{A}, t\in\mathcal{T} \tag{2.9}$$

$$I_{a,t}^{glo} = I_{a,t-1}^{glo} + P_{a,t} - \sum_{r\in\mathcal{R}} O_{r,a,t} \qquad \forall\, a\in\mathcal{A}, t\in\mathcal{T} \tag{2.10}$$

$$I_{a,0}^{glo} = i_{a,0}^{glo} \qquad \forall\, a\in\mathcal{A} \tag{2.11}$$

$$I_{a,T}^{glo} \geq i_{a,T}^{glo} \qquad \forall\, a\in\mathcal{A} \tag{2.12}$$

$$I_{r,a,t}^{reg} = I_{r,a,t-1}^{reg} + O_{r,a,t-l_{r,a}} - min\{d_{r,a,t}, v_{r,a,t}\} \qquad \forall\, r\in\mathcal{R}, a\in\mathcal{A}, t\in\mathcal{T} \tag{2.13}$$

$$I_{r,a,0}^{reg} = i_{r,a,0}^{reg} \qquad \forall\, r\in\mathcal{R}, a\in\mathcal{A} \tag{2.14}$$

$$I_{r,a,t}^{reg} \geq \sum_{m=1}^{\lambda_{r,a}} v_{r,a,t+m} \qquad \forall\, r\in\mathcal{R}, a\in\mathcal{A}, t\in\mathcal{T} \tag{2.15}$$

$$I_{r,a,T}^{reg} \geq i_{r,a,T}^{reg} \qquad \forall\, r\in\mathcal{R}, a\in\mathcal{A} \tag{2.16}$$

$$P_{a,t} \geq 0 \qquad \forall\, a\in\mathcal{A}, t\in\mathcal{T} \tag{2.17}$$

$$I_{a,t}^{glo} \geq 0 \qquad \forall\, a\in\mathcal{A}, t\in\mathcal{T} \tag{2.18}$$

$$O_{r,a,t} \geq 0 \qquad \forall\, r\in\mathcal{R}, a\in\mathcal{A}, t\in\mathcal{T} \tag{2.19}$$

By minimizing the objective function (2.8), we obtain the minimal total inventory investment linked to the supply of the volume guarantee $\vec{v}$ for given, deterministic demands.

Constraints (2.9) through (2.12) reflect the global production level. Constraints (2.9) ensure that AI production does not exceed capacity for any AI and period. Constraints (2.10) and (2.11) capture the inventory balance at central AI production sites, given available starting inventories of $i_{a,0}^{glo}$. Constraints (2.12) ensure that at least $i_{a,T}^{glo}$ units of AI a are available at central storage facilities at the end of the planning horizon to ensure sufficient inventories for future periods beyond T.

At the regional level, constraints (2.13) and (2.14) capture regional stock levels while considering regional starting inventories of $i_{a,0}^{glo}$. The discussed minimum regional stock requirements are incorporated via constraints (2.15), which ensure that inventory levels at the end of each period t are sufficient to cover the total volume guarantees of the succeeding $\lambda_{r,a}$ months. As for global storage, constraints (2.16) ensure the minimum regional target stocks of $i^{reg}_{r,a,T}$ at the end of the planning horizon. Finally, constraints (2.17) through (2.19) are nonnegativity constraints for AI production, global inventories, and regional ordering decisions, respectively. Note that both $d_{r,a,t}$ and $v_{r,a,t}$ are input parameters to the supply chain sub-problem and the *min*-expressions in (2.13) do thus not violate the linearity assumption of the model.

Robust supply chain model

As stated in Section 2.4.1, customer demand is uncertain. To ensure adherence to the inventory target even under uncertain demands, we reformulate the presented deterministic supply chain model into a robust optimization problem to determine the worst-case inventory investment linked to a given set of volume guarantees ($\vec{v}$) under uncertain demands. To this end, we define the uncertainty set of demand $\mathcal{U}$ as a box bounded by the support of the demand distribution ($supp(d_{r,a,t}) = [d^-_{r,a,t}, d^+_{r,a,t}]$):

$$\mathcal{U} = \{d_{r,a,t} : d_{r,a,t} \in [d^-_{r,a,t},\, d^+_{r,a,t}], \forall\, r \in \mathcal{R}, a \in \mathcal{A}, t \in \mathcal{T}\}. \tag{2.20}$$

Recall that we assume that the SD will not request volumes exceeding the upper bound of the uncertain demand distribution, i.e., $v_{r,a,t} \leq d^+_{r,a,t}\, \forall\, r \in \mathcal{R}, a \in \mathcal{A}, t \in \mathcal{T}$.

To robustify the inventory balance equality constraints (2.10), (2.11), (2.13) and (2.14), we eliminate the inventory state variables $I^{glo}_{a,t}$ and $I^{reg}_{r,a,t}$ and reformulate the inventory balance equations in terms of the cumulative product in- and outflow (comp. Ben-Tal et al.,2009):

$$I^{glo}_{a,t} = i^{glo}_{a,0} + \sum_{t'=1}^{t}(P_{a,t'} - \sum_{r \in \mathcal{R}} O_{r,a,t'}) \quad \forall\, a \in \mathcal{A}, t \in \mathcal{T} \tag{2.21}$$

$$I^{reg}_{r,a,t} = i^{reg}_{r,a,0} + \sum_{t'=1}^{t-l_{r,a}} O_{r,a,t'} - \sum_{t'=1}^{t} v_{r,a,t'} \quad \forall\, r \in \mathcal{R}, a \in \mathcal{A}, t \in \mathcal{T} \tag{2.22}$$

Moreover, we reformulate the objective function as the maximum over all demand scenarios.

Using these ingredients, we then formulate the following robust linear program to obtain the minimum worst-case total inventory investment ($WC(\vec{v})$) associated with a given set of volume guarantees ($\vec{v}$) under the defined uncertainty set:

$$WC(\vec{v}) = \min \quad F \tag{2.23}$$

$$\text{s.t.} \sum_{a\in\mathcal{A}}\sum_{t\in\mathcal{T}}(p_a^{glo}(i_{a,0}^{glo} + \sum_{t'=1}^{t}(P_{a,t'} - \sum_{r\in\mathcal{R}} O_{r,a,t'})) \tag{2.24}$$

$$+ \sum_{r\in\mathcal{R}} p_{r,a}^{reg}(i_{r,a,0}^{reg} + \sum_{t'=1}^{t}(O_{r,a,t'} - min\{d_{r,a,t'}, v_{r,a,t'}\})))$$

$$\leq F \quad \forall\, d_{r,a,t'} \in \mathcal{U}$$

$$P_{a,t} \leq c_a \quad \forall\, a \in \mathcal{A}, t \in \mathcal{T} \tag{2.25}$$

$$i_{a,0}^{glo} + \sum_{t'=1}^{t}(P_{a,t'} - \sum_{r\in\mathcal{R}} O_{r,a,t'}) \geq 0 \quad \forall\, a \in \mathcal{A}, t \in \mathcal{T} \tag{2.26}$$

$$i_{a,0}^{glo} + \sum_{t\in\mathcal{T}}(P_{a,t} - \sum_{r\in\mathcal{R}} O_{r,a,t}) \geq i_{a,T}^{glo} \quad \forall\, a \in \mathcal{A} \tag{2.27}$$

$$i_{r,a,0}^{reg} + \sum_{t'=1}^{t-l_{r,a}} O_{r,a,t'} - \sum_{t'=1}^{t} min\{d_{r,a,t'}, v_{r,a,t'}\} \geq \sum_{m=1}^{\lambda_{r,a}} v_{r,a,t+m} \tag{2.28}$$

$$\forall\, r \in \mathcal{R}, a \in \mathcal{A}, t \in \mathcal{T},\; d_{r,a,t'} \in \mathcal{U}$$

$$i_{r,a,0}^{reg} + \sum_{t'=1}^{t-l_{r,a}} O_{r,a,t'} - \sum_{t'=1}^{t} min\{d_{r,a,t'}, v_{r,a,t'}\} \geq i_{r,a,T}^{reg} \quad \forall\, r \in \mathcal{R}, a \in \mathcal{A},\; d_{r,a,t'} \in \mathcal{U} \tag{2.29}$$

$$P_{a,t} \quad \geq 0 \quad \forall\, a \in \mathcal{A}, t \in \mathcal{T} \tag{2.30}$$

$$O_{r,a,t} \geq 0 \quad \forall\, r \in \mathcal{R}, a \in \mathcal{A}, t \in \mathcal{T} \tag{2.31}$$

In this approach, the introduction of demand uncertainty leads to a semi-infinite problem that is computationally intractable. However, we can reformulate the above model as a computationally tractable one by replacing the for all ($\forall$) quantifiers on the uncertainty set with the worst-case realization of demand, i.e., the realization of demand which tightens the respective block of constraints the most (Gorissen et al. 2015).

In addition, we use $min\{d_{r,a,t}^{-}, v_{r,a,t}\}$ to express the minimum quantity sold of AI a in region r and period t which equals the lower limit of uncertain demand if at least the certain component is supplied and the agreed volume guarantee otherwise.

We obtain the following tractable reformulation:

$$WC(\vec{v}) = \min \quad F \tag{2.32}$$

$$\text{s.t.} \sum_{a\in\mathcal{A}} \sum_{t\in\mathcal{T}} (p_a^{glo}(i_{a,0}^{glo} + \sum_{t'=1}^{t}(P_{a,t'} - \sum_{r\in\mathcal{R}} O_{r,a,t'})) \tag{2.33}$$

$$+ \sum_{r\in\mathcal{R}} p_{r,a}^{reg}(i_{r,a,0}^{reg} + \sum_{t'=1}^{t}(O_{r,a,t'} - min\{d_{r,a,t'}^{-}, v_{r,a,t'}\}))) \leq F$$

$$P_{a,t} \leq c_a \quad \forall\, a \in \mathcal{A}, t \in \mathcal{T} \tag{2.34}$$

$$i_{a,0}^{glo} + \sum_{t'=1}^{t}(P_{a,t'} - \sum_{r\in\mathcal{R}} O_{r,a,t'}) \geq 0 \quad \forall\, a \in \mathcal{A}, t \in \mathcal{T} \tag{2.35}$$

$$i_{a,0}^{glo} + \sum_{t\in\mathcal{T}}(P_{a,t} - \sum_{r\in\mathcal{R}} O_{r,a,t}) \geq i_{a,T}^{glo} \quad \forall\, a \in \mathcal{A} \tag{2.36}$$

$$i_{r,a,0}^{reg} + \sum_{t'=1}^{t-l_{r,a}} O_{r,a,t'} - \sum_{t'=1}^{t} v_{r,a,t'} \geq \sum_{m=1}^{\lambda_{r,a}} v_{r,a,t+m} \quad \forall\, r \in \mathcal{R}, a \in \mathcal{A}, t \in \mathcal{T} \tag{2.37}$$

$$i_{r,a,0}^{reg} + \sum_{t'=1}^{t-l_{r,a}} O_{r,a,t'} - \sum_{t'=1}^{t} v_{r,a,t'} \geq i_{r,a,T}^{reg} \quad \forall\, r \in \mathcal{R}, a \in \mathcal{A} \tag{2.38}$$

$$P_{a,t} \quad \geq 0 \quad \forall\, a \in \mathcal{A}, t \in \mathcal{T} \tag{2.39}$$

$$O_{r,a,t} \geq 0 \quad \forall\, r \in \mathcal{R}, a \in \mathcal{A}, t \in \mathcal{T} \tag{2.40}$$

Affinely adjustable robust supply chain model

The presented robust supply chain model assumes that all decisions required to secure the supply guarantees are static, i.e., here-and-now decisions that must be fixed during the budgeting process, prior to any realization of demand uncertainty. In reality, the supply processes can commonly be adjusted, to some degree, in response to market signals that materialize as the planning horizon elapses. Ignoring this flexibility may lead to an overly conservative worst-case assessment in a robust optimization formulation (Ben-Tal et al. 2004). For the budgeting problem at hand, this translates into overstated worst-case inventories and thus into under-utilizing the allowable inventory investment budget, at the expense of avoidable lost margins.

Anticipating and appropriately capturing the ability of the supply chain organization to adapt its supply plan in response to observed market signals is therefore crucial for realistically assessing worst-case inventories of potential sets of volume guarantees. One way to capture the dynamic nature of supply chain decisions is to reformulate the static robust model as a stochastic dynamic program (SDP) with a min-max-structured objective function. Solving an SDP recursively for the T periods of the planning horizon, however, imposes substantial computational challenges and is not a viable course of action for the problem at hand.

As an alternative approach to incorporating dynamic decisions into robust optimization, Ben-Tal et al. (2004) introduce the concept of adjustable robust optimization (ARO). Here, decision variables are not defined as static here-and-now decisions but as decision policies contingent on realized uncertainties while still providing the same robust guarantees. Ben-Tal et al. (2004) show that while solutions determined through an ARO can be substantially less conservative than their static counterparts, determining general decision policies is again computationally intractable. This challenges the applicability of the concept to real-world problems. The authors, therefore, suggest approximating decision policies through affine functions of the realized uncertainties. Similar affine approximation strategies are also applied to other modeling techniques used for uncertain multi-stage problems to address the "curse of dimensionality". Chen (2007) and Chen et al. (2008) use affine decision rules to formulate a tractable multi-period stochastic program, de Farias and Roy (2003) use a linear approximation of the cost-to-go function for an approximate dynamic programming approach, and Shapiro (2011) discusses the conceptual link between adjustable robust optimization, dynamic, and stochastic programming in general.

In an inventory planning context, treating decisions as contingent on demand realizations is intuitively appealing. For example, under a simple base-stock policy with backordering, replenishment orders equal demand - a special case of an affine function of demand. The AARO approach generalizes this policy by allowing other affine functions. If unmet demand is lost, rather than backordered, replenishment orders under a base-stock policy are equal to and thus an affine function of realized sales, rather than demand.

To apply AARO to our worst-case inventory assessment problem, we replace the decision variables in the static RO model (2.23) - (2.31) by decision policies. Since in our setting unmet demand is lost, we consider policies that are affine functions of realized sales ($min\{d_{r,a,t}, v_{r,a,t}\}$), analogous to the base-stock example sketched above. Specifically, we consider the following affinely adjustable decision policies for the ordering ($O_{r,a,t}$) and production ($P_{a,t}$) decisions:

$$O_{r,a,t} = O^0_{r,a,t} + \sum_{\tau=1}^{t-\delta} O^\tau_{r,a,t} min\{d_{r,a,\tau}, v_{r,a,\tau}\} \qquad \forall r \in \mathcal{R}, a \in \mathcal{A}, t \in \mathcal{T} \qquad (2.41)$$

$$P_{a,t} = P^0_{a,t} + \sum_{r \in \mathcal{R}} \sum_{\tau=1}^{t-\delta} P^\tau_{r,a,t} min\{d_{r,a,\tau}, v_{r,a,\tau}\} \qquad \forall r \in \mathcal{R}, a \in \mathcal{A}, t \in \mathcal{T} \qquad (2.42)$$

In this formulation, $O^\tau_{r,a,t}$ and $P^\tau_{r,a,t}$ are decision variables that reflect the per unit impact of realized sales of AI a in region r in period τ on the ordering decision $O_{r,a,t}$, in region r for AI a in period t and the production decision $P_{a,t}$ of AI a in period t, respectively. The decision variables $O^0_{r,a,t}$ and $P^0_{a,t}$ are corresponding intercepts. To illustrate the approach, we note that by setting $O^0_{r,a,1} = S, \quad O^{t-1}_{r,a,t} = 1$ for $t \geq 2$ and dropping all other coefficients and intercepts, we obtain a base-stock policy with base-stock level S for AI a in region r.

As pointed out previously, the AARO formulation (2.41) is more general since it allows more general affine dependencies. However, note that while we allow production quantities to depend on sales in all regions, we consider ordering decisions to depend on sales in that same region only. Finally, we use the parameter $\delta \geq 0$ to capture potential information delays that restrict the timely availability of market signals. To this end, we only allow ordering and production decisions to respond to sales realized at least δ periods before.

By replacing the ordering and production variables in (2.23) - (2.31) using the affine expressions (2.41) and (2.42), we obtain an AARO formulation of the worst-case inventory investment assessment problem that factors in the SCD's ability to adjust planned production and order quantities as actual sales unfold over time. However, as for the initial static RO, the formulation obtained in this way is computationally intractable. Yet we can again rewrite the model in a tractable form by identifying the most constraining demand realizations in each of the constraints.

In the following, we first illustrate this reformulation for individual constraints, before providing the full tractable AARO model. We closely follow the approach of Ben-Tal et al. (2004).

Consider the robust non-negativity constraints (2.31) of the ordering decisions for all regions, AIs and months. Using (2.41), we obtain the affinely adjustable version (2.43) of these constraints. Which demand scenario is the most constraining depends on the sign of $O^{\tau}_{r,a,t}$. For a positive coefficient, it is $d^{-}_{r,a,\tau}$, corresponding with sales of $min\{d^{-}_{r,a,\tau}, v_{r,a,\tau}\}$, whereas for a negative coefficient it is $d^{+}_{r,a,\tau}$, corresponding with sales of $v_{r,a,\tau}$, given our assumption that supply guarantees never exceed maximum possible demand. Using this observation, and introducing a new decision variable $\bar{O}^{\tau}_{r,a,t}$ capturing the negative part of $O^{\tau}_{r,a,t}$ we can rewrite the considered non-negativity constraint as follows.

$$O^0_{r,a,t} + \sum_{\tau=1}^{t-\delta} O^{\tau}_{r,a,t} min\{d_{r,a,\tau}, v_{r,a,\tau}\} \geq 0 \quad \forall\, r \in \mathcal{R}, a \in \mathcal{A}, t \in \mathcal{T},\ d_{r,a,\tau} \in \mathcal{U} \tag{2.43}$$

$$\Leftrightarrow$$

$$O^0_{r,a,t} + \sum_{\tau=1}^{t-\delta} O^{\tau}_{r,a,t} min\{d^{-}_{r,a,\tau}, v_{r,a,\tau}\} + \sum_{\tau=1:O^{\tau}_{r,a,t}<0}^{t-\delta} O^{\tau}_{r,a,t}(v_{r,a,\tau} - min\{d^{-}_{r,a,\tau}, v_{r,a,\tau}\}) \tag{2.44}$$
$$\geq 0 \quad \forall\, r \in \mathcal{R}, a \in \mathcal{A}, t \in \mathcal{T}$$

$$\Leftrightarrow$$

$$O^0_{r,a,t} + \sum_{\tau=1}^{t-\delta} O^{\tau}_{r,a,t} min\{d^{-}_{r,a,\tau}, v_{r,a,\tau}\} + \sum_{\tau=1}^{t-\delta} \bar{O}^{\tau}_{r,a,t}(v_{r,a,\tau} - min\{d^{-}_{r,a,\tau}, v_{r,a,\tau}\}) \geq 0 \tag{2.45}$$
$$\forall\, r \in \mathcal{R}, a \in \mathcal{A}, t \in \mathcal{T}$$
$$\bar{O}^{\tau}_{r,a,t} \leq 0 \qquad \forall r \in \mathcal{R}, a \in \mathcal{A}, \tau \in \mathcal{T}, t \in \mathcal{T}$$
$$\bar{O}^{\tau}_{r,a,t} \leq O^{\tau}_{r,a,t} \quad \forall r \in \mathcal{R}, a \in \mathcal{A}, \tau \in \mathcal{T}, t \in \mathcal{T}$$

We can use a similar approach to reformulate $\leq$-constraints, as illustrated here for the capacity constraints (2.34).

$$P^0_{a,t} + \sum_{r\in\mathcal{R}} \sum_{\tau=1}^{t-\delta} P^{\tau}_{r,a,t} min\{d_{r,a,\tau}, v_{r,a,\tau}\} \leq c_a \quad \forall\, a \in \mathcal{A}, t \in \mathcal{T}, d_{r,a,\tau} \in \mathcal{U} \tag{2.46}$$

$$\Leftrightarrow$$

$$P^0_{a,t} + \sum_{r \in \mathcal{R}} \sum_{\tau=1}^{t-\delta} P^{\tau}_{r,a,t} min\{d^-_{r,a,\tau}, v_{r,a,\tau}\} + \sum_{r \in \mathcal{R}} \sum_{\tau=1: P^{\tau}_{r,a,t}>0}^{t-\delta} P^{\tau}_{r,a,t}(v_{r,a,\tau} - min\{d^-_{r,a,\tau}, v_{r,a,\tau}\}) \leq c_a \quad \forall\, a \in \mathcal{A}, t \in \mathcal{T} \tag{2.47}$$

$$\Leftrightarrow$$

$$P^0_{a,t} + \sum_{r \in \mathcal{R}} \sum_{\tau=1}^{t-\delta} P^{\tau}_{r,a,t} v_{r,a,\tau} + \sum_{r \in \mathcal{R}} \sum_{\tau=1: P^{\tau}_{r,a,t}<0}^{t-\delta} P^{\tau}_{r,a,t}(v_{r,a,\tau} - min\{d^-_{r,a,\tau}, v_{r,a,\tau}\}) \leq c_a \quad \forall\, a \in \mathcal{A}, t \in \mathcal{T} \tag{2.48}$$

$$\Leftrightarrow$$

$$P^0_{a,t} + \sum_{r \in \mathcal{R}} \sum_{\tau=1}^{t-\delta} P^{\tau}_{r,a,t} v_{r,a,\tau} - \sum_{r \in \mathcal{R}} \sum_{\tau=1}^{t-\delta} \bar{P}^{\tau}_{r,a,t}(v_{r,a,\tau} - min\{d^-_{r,a,\tau}, v_{r,a,\tau}\}) \leq c_a \quad \forall\, a \in \mathcal{A}, t \in \mathcal{T} \tag{2.49}$$

$$\bar{P}^{\tau}_{r,a,t} \leq 0 \qquad \forall r \in \mathcal{R}, a \in \mathcal{A}, \tau \in \mathcal{T}, t \in \mathcal{T}$$

$$\bar{P}^{\tau}_{r,a,t} \leq P^{\tau}_{r,a,t} \quad \forall r \in \mathcal{R}, a \in \mathcal{A}, \tau \in \mathcal{T}, t \in \mathcal{T}$$

Repeating these steps for all constraints, we can convert the complete static RO into a tractable formulation of its AARO counterpart. For notational convenience and following Ben-Tal et al. (2004), we introduce additional variables

$$I^{glo,\tau}_{r,a,t} \equiv \sum_{t'=1}^{t} (P^{\tau}_{r,a,t'} - O^{\tau}_{r,a,t'}) \quad \forall r \in \mathcal{R}, a \in \mathcal{A}, \tau \in \mathcal{T}, t \in \mathcal{T} \tag{2.50}$$

$$I^{reg,\tau}_{r,a,t} \equiv \sum_{t'=1}^{t-l_{r,a}} O^{\tau}_{r,a,t'} - 1 \quad \forall r \in \mathcal{R}, a \in \mathcal{A}, \tau \in \mathcal{T}, t \in \mathcal{T} \tag{2.51}$$

$$F^{\tau}_{r,a} \equiv \sum_{t \in \mathcal{T}} (p^{glo}_a\, I^{glo,\tau}_{r,a,t} + p^{reg}_{r,a}(I^{reg,\tau}_{r,a,t} + \sum_{i=0: t+i \leq T}^{l_{r,a}-1} O^{\tau}_{r,a,t})) \quad \forall r \in \mathcal{R}, a \in \mathcal{A}, \tau \in \mathcal{T} \tag{2.52}$$

which capture the per unit impact of realized sales on global and regional stock levels and on the objective value. Using this notation, we obtain the following final AARO formulation:

$$WC(\vec{v}) = \min \quad F \tag{2.53}$$

s.t.

<u>*Objective function:*</u>

$$\begin{aligned}(\sum_{a\in\mathcal{A}}\sum_{t\in\mathcal{T}}(p^{glo}_a i^{glo}_{a,0} + \sum_{r\in\mathcal{R}} p^{reg}_{r,a} i^{reg}_{r,a,0} + \sum_{t'=1}^{t}(p^{glo}_a(P^0_{a,t'} - \sum_{r\in\mathcal{R}} O^0_{r,a,t'}) + p^{reg}_{r,a}\sum_{r\in\mathcal{R}} O^0_{r,a,t'})) \\ + \sum_{r\in\mathcal{R}}\sum_{a\in\mathcal{A}}\sum_{\tau\in\mathcal{T}}(F^\tau_{r,a} v_{r,a,\tau} - \bar{F}^\tau_{r,a}(v_{r,a,\tau} - min\{d^-_{r,a,\tau}, v_{r,a,\tau}\}))) \leq F\end{aligned} \tag{2.54}$$

$$\sum_{t\in\mathcal{T}}(p^{glo}_a I^{glo,\tau}_{r,a,t} + p^{reg}_{r,a}(I^{reg,\tau}_{r,a,t} + \sum_{i=0:t+i\leq T}^{l_{r,a}-1} O^\tau_{r,a,t})) = F^\tau_{r,a} \quad \forall r\in\mathcal{R}, a\in\mathcal{A}, \tau\in\mathcal{T} \tag{2.55}$$

$$\bar{F}^\tau_{r,a} \leq 0 \qquad \forall r\in\mathcal{R}, a\in\mathcal{A}, \tau\in\mathcal{T} \tag{2.56}$$

$$\bar{F}^\tau_{r,a} \leq F^\tau_{r,a} \quad \forall r\in\mathcal{R}, a\in\mathcal{A}, \tau\in\mathcal{T} \tag{2.57}$$

<u>*Global AI production constraints:*</u>

$$\begin{aligned}P^0_{a,t} + \sum_{r\in\mathcal{R}}\sum_{\tau=1}^{t-\delta} P^\tau_{r,a,t} min\{d^-_{r,a,\tau}, v_{r,a,\tau}\} + \sum_{r\in\mathcal{R}}\sum_{\tau=1}^{t-\delta} \bar{P}^\tau_{r,a,t}(v_{r,a,\tau} - min\{d^-_{r,a,\tau}, v_{r,a,\tau}\}) \\ \geq 0 \quad \forall a\in\mathcal{A}, t\in\mathcal{T}\end{aligned} \tag{2.58}$$

$$\begin{aligned}P^0_{a,t} + \sum_{r\in\mathcal{R}}\sum_{\tau=1}^{t-\delta} P^\tau_{r,a,t} v_{r,a,\tau} - \sum_{r\in\mathcal{R}}\sum_{\tau=1}^{t-\delta} \bar{P}^\tau_{r,a,t}(v_{r,a,\tau} - min\{d^-_{r,a,\tau}, v_{r,a,\tau}\}) \\ \leq c_a \quad \forall a\in\mathcal{A}, t\in\mathcal{T}\end{aligned} \tag{2.59}$$

$$\sum_{t'=1}^{t}(P^\tau_{r,a,t'} - O^\tau_{r,a,t'}) = I^{glo,\tau}_{r,a,t} \quad \forall r\in\mathcal{R}, a\in\mathcal{A}, \tau\in\mathcal{T}, t\in\mathcal{T} \tag{2.60}$$

$$\begin{aligned}i^{glo}_{a,0} + \sum_{t'=1}^{t}(P^0_{a,t'} - \sum_{r\in\mathcal{R}} O^0_{r,a,t'}) + \sum_{r\in\mathcal{R}}\sum_{\tau=1}^{t} I^{glo,\tau}_{r,a,t} min\{d^-_{r,a,\tau}, v_{r,a,\tau}\} \\ + \sum_{r\in\mathcal{R}}\sum_{\tau=1}^{t} \bar{I}^{glo,\tau}_{r,a,t}(v_{r,a,\tau} - min\{d^-_{r,a,\tau}, v_{r,a,\tau}\}) \geq 0 \qquad \forall a\in\mathcal{A}, t\in\mathcal{T}\end{aligned} \tag{2.61}$$

$$\begin{aligned}i^{glo}_{a,0} + \sum_{t\in\mathcal{T}}(P^0_{a,t} - \sum_{r\in\mathcal{R}} O^0_{r,a,t}) + \sum_{r\in\mathcal{R}}\sum_{\tau\in\mathcal{T}} I^{glo,\tau}_{r,a,T} min\{d^-_{r,a,\tau}, v_{r,a,\tau}\} \\ + \sum_{r\in\mathcal{R}}\sum_{\tau\in\mathcal{T}} \bar{I}^{glo,\tau}_{r,a,T}(v_{r,a,\tau} - min\{d^-_{r,a,\tau}, v_{r,a,\tau}\}) \geq i^{glo}_{a,T} \quad \forall a\in\mathcal{A}\end{aligned} \tag{2.62}$$

$$\bar{P}^{\tau}_{r,a,t} \leq 0 \qquad \forall r \in \mathcal{R}, a \in \mathcal{A}, \tau \in \mathcal{T}, t \in \mathcal{T} \tag{2.63}$$

$$\bar{P}^{\tau}_{r,a,t} \leq P^{\tau}_{r,a,t} \qquad \forall r \in \mathcal{R}, a \in \mathcal{A}, \tau \in \mathcal{T}, t \in \mathcal{T} \tag{2.64}$$

$$\bar{I}^{glo,\tau}_{r,a,t} \leq 0 \qquad \forall r \in \mathcal{R}, a \in \mathcal{A}, \tau \in \mathcal{T}, t \in \mathcal{T} \tag{2.65}$$

$$\bar{I}^{glo,\tau}_{r,a,t} \leq I^{glo,\tau}_{r,a,t} \qquad \forall r \in \mathcal{R}, a \in \mathcal{A}, \tau \in \mathcal{T}, t \in \mathcal{T} \tag{2.66}$$

$$P^{\tau}_{r,a,t} = 0 \qquad \forall r \in \mathcal{R}, a \in \mathcal{A}, \tau \in \mathcal{T}, t \in \mathcal{T}, \tau + \delta > t \tag{2.67}$$

<u>*Regional constraints:*</u>

$$O^{0}_{r,a,t} + \sum_{\tau=1}^{t-\delta} O^{\tau}_{r,a,t} min\{d^{-}_{r,a,\tau}, v_{r,a,\tau}\} + \sum_{\tau=1}^{t-\delta} \bar{O}^{\tau}_{r,a,t}(v_{r,a,\tau} - min\{d^{-}_{r,a,\tau}, v_{r,a,\tau}\}) \geq 0 \qquad \forall \quad r \in \mathcal{R}, a \in \mathcal{A}, t \in \mathcal{T} \tag{2.68}$$

$$\begin{aligned} i^{reg}_{r,a,0} + \sum_{t'=1}^{t-l_{r,a}} O^{0}_{r,a,t'} &+ \sum_{\tau=1}^{t} I^{reg,\tau}_{r,a,t} min\{d^{-}_{r,a,\tau}, v_{r,a,\tau}\} \\ &+ \sum_{\tau=1}^{t} \bar{I}^{reg,\tau}_{r,a,t}(v_{r,a,\tau} - min\{d^{-}_{r,a,\tau}, v_{r,a,\tau}\}) \\ &\geq \sum_{m=1}^{\lambda_{r,a}} v_{r,a,t+m} \quad \forall r \in \mathcal{R}, a \in \mathcal{A}, t \in \mathcal{T} \end{aligned} \tag{2.69}$$

$$\begin{aligned} i^{reg}_{r,a,0} + \sum_{t=1}^{T-l_{r,a}} O^{0}_{r,a,t} &+ \sum_{\tau \in \mathcal{T}} I^{reg,\tau}_{r,a,T} min\{d^{-}_{r,a,\tau}, v_{r,a,\tau}\} \\ &+ \sum_{\tau \in \mathcal{T}} \bar{I}^{reg,\tau}_{r,a,T}(v_{r,a,\tau} - min\{d^{-}_{r,a,\tau}, v_{r,a,\tau}\}) \\ &\geq i^{reg}_{r,a,T} \quad \forall r \in \mathcal{R}, a \in \mathcal{A} \end{aligned} \tag{2.70}$$

$$\sum_{t'=1}^{t-l_{r,a}} O^{\tau}_{r,a,t'} - 1 = I^{reg,\tau}_{r,a,t} \quad \forall r \in \mathcal{R}, a \in \mathcal{A}, \tau \in \mathcal{T}, t \in \mathcal{T} \tag{2.71}$$

$$\bar{O}^{\tau}_{r,a,t} \leq 0 \qquad \forall r \in \mathcal{R}, a \in \mathcal{A}, \tau \in \mathcal{T}, t \in \mathcal{T} \tag{2.72}$$

$$\bar{O}^{\tau}_{r,a,t} \leq O^{\tau}_{r,a,t} \qquad \forall r \in \mathcal{R}, a \in \mathcal{A}, \tau \in \mathcal{T}, t \in \mathcal{T} \tag{2.73}$$

$$\bar{I}^{reg,\tau}_{r,a,t} \leq 0 \qquad \forall r \in \mathcal{R}, a \in \mathcal{A}, \tau \in \mathcal{T}, t \in \mathcal{T} \tag{2.74}$$

$$\bar{I}^{reg,\tau}_{r,a,t} \leq I^{reg,\tau}_{r,a,t} \qquad \forall r \in \mathcal{R}, a \in \mathcal{A}, \tau \in \mathcal{T}, t \in \mathcal{T} \tag{2.75}$$

$$O^{\tau}_{r,a,t} = 0 \qquad \forall r \in \mathcal{R}, a \in \mathcal{A}, \tau \in \mathcal{T}, t \in \mathcal{T}, \tau + \delta > t \tag{2.76}$$

Through the presented AARO model, the SCD can obtain an assessment of worst-case inventories of a given set of volume guarantees that considers uncertainty and, in contrast to static RO, the existing flexibility for adjusting relevant decisions over time. As illustrated in Figure 2.1, the SCD's assessment serves as subproblem (2.5) - (2.6) of the budget planning problem.

2.5. Numerical study based on synthetic data

Our goal in this section is to gain insights into the performance and solution structure of our proposed modeling approach. To this end, we use a synthetic data set which allows us to systematically vary the input parameters.

2.5.1. Study setup and benchmark strategies

We assess the performance of the guarantees ($\vec{v}^*$) derived with the proposed budget planning model, which we refer to as DG (differentiated guarantees), by comparing them to those of two simple benchmark strategies: fixed factor (FF) and fixed guarantee level (FL).

Under a FF strategy, the SCD guarantees the supply of a fixed fraction f of the demand point forecast for any AI-region-time combination, i.e. $v_{r,a,t} = f \cdot d^{FC}_{r,a,t}$ with some fixed multiplier f. This approach comes closest to the current practice at the case company. Effectively tailoring volume guarantees to different AIs and sales regions is prohibitively complex for a manual approach. Moreover, decision-makers find it difficult to systematically factor in demand uncertainty. The S&OP process at the case company therefore largely uses a fixed fraction of the point forecast and concentrates on calibrating this fixed fraction to match the allowable inventory investment. In our analysis, we chose f by numerically determining the highest multiplier for which the resulting guarantees can still be supplied by the SCD without exceeding the inventory investment target.

The FL rule is a slightly more advanced extension of FF. It again guarantees a fixed fraction, but applies it to the uncertainty range of demand, rather than to a point forecast, i.e. $v_{r,a,t} = d^-_{r,a,t} + g \cdot (d^+_{r,a,t} - d^-_{r,a,t})$. We calibrate the fraction g by numerically deriving the highest g for which the resulting guarantees can still be supplied without exceeding the inventory investment target.

In addition, we compare the performance of the proposed AARO-based budget planning model (DG) to an approach that assesses worst-case inventories based on the static RO model presented in Section 2.4.3. We denote this approach as DG^{static}. This comparison provides insights into the importance of taking supply adjustment flexibility into account during the budget process.

In our numerical study, we consider 18 AIs supplied to three sales regions, leading to 18x3=54 regional AI value chains. This problem size is similar to the one in the real-life case application presented in Section 2.6. Within this setting, we investigate four scenarios that differ in the degree of diversity in parameter values between regions and products. Specifically, we combine high/low diversity of supply parameters with high/low diversity of demand parameters.

In smaller prestudies, we have identified the supply and demand parameters most relevant to the overall problem results, namely, distribution lead time, regional cost of goods, product margins, and level of demand uncertainty. For each of these, we define two sets of relatively homogeneous versus markedly heterogeneous parameter values (see Table 2.1). For a given scenario, we then parameterize each of the 54 regional AI value chains through a different combination of the 3x3x3x2=54 values of $l_{r,a}$, $p_{r,a}^{reg}$, $m_{r,a}$ and $\frac{d_{r,a,t}^{+}-d_{r,a,t}^{-}}{\mathbb{E}(d_{r,a,t})}$ in the respective cells in Table 2.1.

		Scenarios	
Domain	Parameter	Low diversity	High diversity
Supply	$l_{r,a}$	{1, 1, 1}	{0, 1, 2}
	$p_{r,a}^{reg}$	{7, 10, 13}	{3, 10, 17}
Demand	$m_{r,a}$	{7, 10, 13}	{3, 10, 17}
	$\frac{d_{r,a,t}^{+}-d_{r,a,t}^{-}}{\mathbb{E}(d_{r,a,t})}$	{0.36, 0.44}	{0.20, 0.60}

Table 2.1.: Sets of values of relevant supply and demand parameters used to construct different benchmark portfolios with a low or high degree of parameter diversity.

For simplicity and transparency, we keep the other parameters constant across all portfolios and set them to realistic levels observed in the industry, relative to the other parameters: $c_a = 390$, $p_{r,a}^{glo} = 2$, $\lambda_{r,a} = 2$, $\delta = 1$, $T = 24$.

Similarly, we consider a single monthly demand pattern that follows a typical seasonal cycle commonly observed at the case company and shown in Table 2.2. Forecast errors are sampled from a normal, zero-mean symmetrically cropped distribution with a standard deviation such that $\mathbb{E}(d_{r,a,t}) \pm 0.5(d^{+}_{r,a,t} - d^{-}_{r,a,t})$ corresponds to its support and the 95^{th} percentile of the original uncropped normal distribution.

Period	1	2	3	4	5	6	7	8	9	10	11	12
Exp. Demand	4	6	7	10	11	12	12	11	10	7	6	4

Table 2.2.: Seasonal monthly demand pattern used for the synthetic data study. Months 1 to 12 are identical to months 13 to 24.

For each of the four scenarios, we derive the expected lost-margin-minimal guarantees according to the proposed budget planning model.

As the budget planning problem (2.3) - (2.6) is structurally complex, we cannot solve it exactly. Instead, we derive the guarantees for *DG* heuristically through a gradient-based local search. To simplify the search, we follow the practice of the case company to focus the discussions during budgeting on deriving total seasonal guarantee levels per AI and region for the upcoming business year. Following this practice, we drop the time index t in the guarantee vector and assume that guarantee levels for regional AI demands are chosen such that the same share of the uncertain demand component is covered across time:

$$v_{r,a,t}(g_{r,a}) = d^{-}_{r,a,t} + g_{r,a}(d^{+}_{r,a,t} - d^{-}_{r,a,t}) \quad \forall r \in \mathcal{R}, a \in \mathcal{A}, t \in \mathcal{T} \tag{2.77}$$

Details on the solution heuristic are provided in Appendix A.

Finally, we account for a business rule at the case company which requires that at least the certain component of uncertain regional AI sales must be supplied. We therefore require $g_{r,a}$ to be nonnegative $\forall r \in \mathcal{R}, a \in \mathcal{A}$. We discuss the implications of this rule in Section 2.6.

2.5.2. Performance analysis of allocation strategies

We now present the core results of our numerical study, which concern the performance of our proposed S&OP support model. To this end, we report the expected lost margins associated with the set of volume guarantees determined by our proposed model and compare them with the solutions of the benchmark strategies on the set of scenarios defined in Section 2.5.1. In addition to the overall results, we analyze their sensitivity along three dimensions: (1) the level of the inventory investment budget and the level of diversity in the relevant (2) supply and (3) demand parameters within the AI portfolio. Table 2.3 summarizes the results. The percentage values in parentheses indicate the relative difference to the results obtained with the FF strategy.

	DG	FL	FF	DG^{static}
Avg.	934.8 (-22.6%)	1161.1 (-3.8%)	1207.2	1769.5
Inventory target (β):				
L	1719.6 (-21.7%)	2117.1 (-3.6%)	2196.6	2229.8
H	150.0 (-31.1%)	205.1 (-5.8%)	217.8	1309.2
Supply Diversity:				
L	1017.3 (-15.7%)	1161.1 (-3.8%)	1207.2	1884.7
H	852.3 (-29.4%)	1161.0 (-3.8%)	1207.2	1654.2
Demand Diversity:				
L	978.3 (-15.8%)	1160.8 (-0.1%)	1162.4	1834.0
H	891.2 (-28.8%)	1161.4 (-7.2%)	1252.0	1705.0

Table 2.3.: Expected total lost margins based on guarantees determined by different budget models for different inventory targets and portfolio parameter diversities ($\delta = 1$).

Across all considered portfolios and inventory investment budgets, the DG-based volume guarantees lead to an average reduction in the expected total lost margin by 22.6% compared to that of the undifferentiated, single-parameter allocation strategy FF. In contrast, FL barely outperforms FF, with a mere 3.8% reduction in the expected total lost margin, on average. These figures clearly document the benefits of differentiated supply guarantees. A manual approach cannot effectively handle this level of complexity. Our approach demonstrates the power of optimization and substantially reduces expected lost margins, relative to current practice, while respecting the same inventory budget.

The results also underline that it is important for the inventory assessment of supply guarantees to factor in inherent sources of flexibility, i.e. capabilities of the SCD to adjust supply decisions in response to updated demand information. Ignoring the option value of such flexibilities leads to unnecessarily low supply guarantees caused by an overly conservative assessment of worst-case inventories. Due to the observed clear performance benefit of the adjustable model formulation relative to the static one, we henceforth exclude DG^{static} from our analysis .

Naturally, the aforementioned effects depend on key problem parameters. We observe the following relationships. We find that the reduction of expected lost margins from DG relative to those from FF increases from 21.7% for a low inventory budget to 31.1% for a high budget. The observed lower reduction potential for low allowable inventory investments is partly due to the company's business rule, which requires that at least the certain component of demand must be supplied. The inventory needed to satisfy this requirement consumes an increasing share of the inventory budget if the budget decreases, thereby leaving less room for optimization. We discuss this issue in more detail in Section 2.6.3. Note that despite the lower percentage values, the absolute reduction in lost margins is much higher for a low inventory budget than for a high budget. In all cases, the benefits of FL over FF are much lower, amounting to 3.6% and 5.8%, on average, for a low and high inventory budget, respectively.

As for the degree of diversity in the relevant supply and demand parameters, we observe that the relative benefit of the proposed DG strategy over that of the benchmarks is larger for more diverse portfolios. On the supply parameter side, the reduction in expected lost margins increases from 15.7% for low diversity to 29.4% for high diversity. For the demand parameters, we observe a similar increase from 15.8% to 28.8%, respectively. Intuitively, a more heterogeneous portfolio offers more opportunities to differentiate and thereby minimize the impact of inventory budget restrictions on lost margins. Considering the benchmarks, FL again only slightly outperforms FF for both low and high diversity in supply parameters. The level of diversity of the demand parameter values appears to have a stronger impact on the relative performance gap between these methods. While FL clearly outperforms FF in high diversity, both methods perform roughly at par for low demand diversity.

In summary, these results highlight the relevance of supporting the S&OP decision-making process on volume guarantees by a systematic model. The size and complexity of the problem prohibit a manual solution, and resorting to simplistic heuristics is not effective. In addition, differentiating supply guarantees based on relevant supply and demand characteristics is especially important if the underlying product portfolio is strongly heterogeneous.

2.5.3. Solution structure and driver analysis

Besides the performance analysis, we investigate the structure of the solution. To isolate the key drivers for differentiation in the resulting regional AI supply guarantees, we apply our model to the portfolio with high supply and demand parameter diversity and derive the optimal sets of guarantees for low and high inventory targets. To simplify the comparison of different regional AI volume guarantees, we translate the derived guarantee levels ($\vec{v}^*$) into the corresponding expected service levels, expressed as the percentage of uncertain margin potential captured by the guaranteed volumes.

To determine the impact and relative importance of problem parameters both individually and interactively for the derived volume guarantees, we ran a linear regression analysis. The problem parameters $p_{r,a}^{reg}$, $l_{r,a}$, $m_{r,a}$, $\frac{d_{r,a,t}^{+}-d_{r,a,t}^{-}}{\mathbb{E}(d_{r,a,t})}$ are treated as the independent variables, and the expected service levels resulting from the optimized regional AI volume guarantees as the dependent variable of the linear regression. We report the resulting standardized regression coefficients under a low and high inventory investment target in Table 2.4.

	Standardized regression coefficients			
Inventory investment (β)	$p_{r,a}^{reg}$	$l_{r,a}$	$m_{r,a}$	$\frac{d_{r,a,t}^{+}-d_{r,a,t}^{-}}{\mathbb{E}(d_{r,a,t})}$
L	-0.45	-0.09	0.38	-0.64
H	-0.38	-0.19	0.46	-0.26

Table 2.4.: Standardized regression coefficients of the linear regression analysis examining the influence of problem parameters on the expected service levels based on guarantees determined through *DG* under a low and high inventory target ($\delta = 1$).

We observe that our AARO model proposes higher service levels for regional AI supply chains with low inventory costs, short lead times, high margins, and low levels of uncertainty.

None of these directional effects are surprising as such. What nevertheless renders the solution complex is the difference in effect strengths and their interaction. For the current data set, our model differentiates volume guarantees and resulting service levels strongly on the basis of demand uncertainty (-0.64) and regional inventory costs (-0.45) if the inventory investment target is low. For high inventory budgets, the differences in profitability (0.46) become more important while the relative importance of the degree of demand uncertainty decreases.

2.5.4. Impact of information and material flow delays on inventory

In the previous sections, we analysed the performance of the proposed budget planning model with respect to expected lost margins and expected service levels, for a given inventory investment target. In the following, we assess drivers of the worst-case inventory investment associated with a given set of volume guarantees. In particular, we investigate how the inventory investment depends on the level of supply chain responsiveness. We express supply chain responsiveness by two factors, namely, the information delay (δ) and the material flow delay caused by the distribution lead time ($l_{r,a}$).

We analyze the changes in the worst-case inventory investment as we vary both factors between 1 and 3 months. Table 2.5 shows the worst-case inventories if the supply of the upper bound of the uncertain demand distribution is guaranteed ($\vec{v} = \vec{d^{+}}$) for a portfolio with a high diversity in both supply and demand parameters (comp. Table 2.1). In addition to the AARO results, the table includes the static RO outcome, which can be interpreted as a special case with an information delay of at least T months.

		$l_{r,a}$		
WC Inventory	δ	1	2	3
AARO	1	3767	5295	6796
	2	4274	5813	7318
	3	4776	6314	7806
RO	T	7179	8707	10197

Table 2.5.: Worst-case inventories based on AARO models with different information delays (δ) and a non-adjustable RO benchmark for different lead times ($l_{r,a}$).

From Table 2.5, we find that while both forms of delays increase the required worst-case inventory investments, the overall impact of distribution lead times is higher. Compared to a worst-case inventory investment of 3767 for an information delay and lead time of one month, worst-case inventories increase by 80% to 6796 when raising the distribution lead time to three months. In contrast, an information delay increase to 3 months lets the inventory investment grow by only 27% to 4776. The strong impact of distribution lead times on inventory levels can be explained by two reasons. First, an increase in transit times directly causes an increase in in-transit stocks. Second, regional orders need to be placed increasingly in advance of the peak sales months, which limits in-season order adjustment flexibility.

Despite its strong impact, reducing transportation times usually comes at a significant cost. It requires either a shift in the mode of transport from ship or truck to air freight or a redesign of the company's production network. Neither may be economically feasible or achievable in the short run. A reduction in information lead times across and within functions, on the other hand, may be achievable by highlighting the value of introducing collaborative processes and information systems to support information sharing and a rapid adjustment of production and distribution decisions throughout the organization in response to recent market signals (Lapide 2004).

Another important observation from Table 2.5 is that it is crucial for an appropriate inventory investment assessment to capture the existing level of supply chain flexibility. For a distribution lead time of one month, a static RO overstates the worst-case inventory investment derived by a corresponding AARO with an information delay of one month by 91% (7179 vs. 3767). Incorporating the ability of the supply chain to adjust decisions over time, in response to unfolding uncertainties is thus key to avoiding an overly conservative allocation of volume guarantees and to preventing unnecessary lost margins.

The qualitative conclusions concerning the impact of information and material flow delays on inventory remain valid for portfolios with less heterogeneous problem parameters.

2.6. Application to real-world data of the case company

In this section, we complement the general findings from the synthetic data analysis by quantifying the improvement potential of our proposed model in a real-world context.

2.6.1. Data collection and scope of analysis

For the performance assessment of the proposed approach in practice, we use real-world data to derive volume guarantees for the 15 most business-relevant AIs of the case company. Each of the AIs is centrally synthesized and stored before being distributed and sold in up to four regions. For each of the AIs, we collect the relevant production capacity, distribution lead times, inventory costs at global and regional storage facilities, and the relevant AI-equivalent sales forecasts for the next 24 months. Moreover, we assess demand uncertainty based on historical forecast errors. To obtain robust distributional information under limited data availability per AI, we pool historical forecast errors of similar AIs within the company's portfolio, i.e., AIs of the same product type with similar sales patterns sold in the same region. We assume that the distribution of relative forecast errors is constant over time and obtain the box uncertainty set based on the distributional bounds. For the assessment of the expected lost margins for a given volume guarantee, we sample from the pooled distribution. To protect the confidentiality interests of the case company, we anonymize the input data and normalize the presented model outputs.

2.6.2. Performance analysis of allocation strategies

In Table 2.6, we compare the resulting expected total lost margin for volume guarantees derived through DG, FL, and FF for low, medium, and high maximum inventory investments (β) and three levels of information delay (δ), ranging between one and three months.

δ	β	DG	FL	FF
1	L	637.3 (-18.0%)	759.9 (-2.2%)	777.0
	M	189.6 (-29.0%)	279.1 (4.5%)	267.1
	H	37.7 (-31.7%)	50.4 (-8.6%)	55.2
2	L	679.7 (-16.1%)	783.9 (-3.2%)	810.2
	M	249.3 (-26.7%)	352.0 (3.6%)	339.9
	H	74.9 (-26.6%)	104.2 (2.2%)	102.0
3	L	710.6 (-15.2%)	802.4 (-4.3%)	838.3
	M	298.2 (-24.9%)	407.9 (2.7%)	397.2
	H	107.6 (-27.2%)	154.8 (4.7%)	147.8

Table 2.6.: Expected lost margins based on guarantees determined by different budget models for different inventory investment targets (β) and information delays (δ).

DG again outperforms the simple, single parameter benchmark strategies, with a reduction in expected lost margins between 15% and 32% compared to that of FF. This range is similar to the one derived for the synthetic data. Additionally, the relative improvement potential of DG over FF again increases in β. The explanation is the same as that discussed in Section 2.5.2.

Considering the reduction potential of the expected lost margin in absolute terms, we observe that it is the largest for tight inventory investment limits. This highlights the importance of agreeing on differentiated volume guarantees during the budgeting process. An increase in the information delay has a negative effect on the lost margin reduction potential, as expected. However, differentiated guarantees remain highly superior to undifferentiated guarantees.

The comparison of the relative performance of the benchmark strategies FL and FF shows a different picture than for the synthetic data. Whereas FL at least marginally outperforms FF for all instances in the synthetic data set, as it links the provision of volume guarantees to the degree of demand uncertainty in an otherwise symmetrically constructed AI portfolio, the results for the real-world data are mixed. None of the simple, single-parameter strategies provides consistently better results than the other.

Overall, the numerical results obtained for the real-world data of the case company's budget planning problem demonstrate a considerable improvement potential of our proposed model over the current planning practice.

2.6.3. Solution structure analysis

In this section, we investigate what factors drive the guarantee allocation for the given real-world data. To this end and in line with the analysis conducted in Section 2.5.3 , Table 2.7 displays the resulting standardized regression coefficients from the linear regression of the expected service levels of volume guarantees determined by DG (dependent variable) on the problem parameters (independent variables) under a low, medium and high inventory investment target.

	Standardized regression coefficients			
Inventory investment (β)	$p^{reg}_{r,a}$	$l_{r,a}$	$m_{r,a}$	$\frac{d^+_{r,a,t}-d^-_{r,a,t}}{\mathbb{E}(d_{r,a,t})}$
L	-0.55	-0.29	0.74	-0.53
M	-0.49	-0.23	0.75	-0.13
H	-0.52	-0.25	0.75	-0.01

Table 2.7.: Standardized regression coefficients of the linear regression analysis examining the influence of problem parameters on the expected service levels based on guarantees determined through DG under a low, medium, and high inventory target ($\delta = 1$).

The signs of all regression coefficients are identical to those in the synthetic data study. This confirms the intuition that the AARO-based budget process prioritizes regional AI supply chains with low inventory costs, short lead times, high margins, and low levels of uncertainty. In addition, an increasing allowable inventory investment again reduces the relative importance of demand uncertainty as a differentiating factor. Different from the synthetic data study, the relative importance of the other three considered factors remains rather stable across different inventory investment levels, with their ranking corresponding to that under high inventory allowance in the synthetic study. We attribute the observed differences between the results of both numerical studies to structural differences in the underlying data sets. Recall that the synthetic data set uses a full factorial combination of the problem parameters listed in Table 2.1. This implies that parameter values are uncorrelated. In the real-world data, in contrast, we found the regional cost of goods and product margin to be positively correlated, which leads to a stable importance of both parameters across the three studied inventory investment levels. Furthermore, the variance in distribution lead times of regional AI supply chains is higher in the real-world data than in the synthetic data set. Consequently, distribution lead time emerges as a more influential differentiation factor, even under scenarios with low inventory investment targets.

All in all, the changes in the relative importance of the different market characteristics underscore the complexity of the budget planning problem and the need for a systematic case-by-case analysis. Simple fixed priority rules are clearly insufficient.

2.7. Conclusion

In this paper, we addressed the cross-functional decision problem faced by an agrochemical supplier as part of the annual budget planning cycle. Through the budgeting process, sales and supply chain activities are aligned by agreeing on a set of volume guarantees to be supplied throughout the planning horizon while adhering to a maximum allowable inventory investment imposed by the business unit head. We developed an optimization-based decision-support tool to obtain these volume guarantees. A key contribution of our model formulation is that it captures the conflicting incentives and asymmetric decision-making power of the parties involved.

In two numerical studies based on synthetic as well as real-world data, we analyzed key drivers for the obtained solutions, and we explored the performance of our proposed model compared to that of the simple heuristics currently in place at the case company. We found that a considerable lost margin reduction can be achieved by allocating more volume guarantees toward AIs suppliable under low lead times and with low per unit inventory costs, high per unit margins, and low demand uncertainty. For the real-world data set, the achievable expected lost margin reduction amounts to 32%.

Our work provides several insights that extend beyond the specific setting of the case company. Much of the existing literature on decision support for S&OP assumes a centralized decision-maker optimizing a single objective function. Such model formulations ignore real-world complexities due to conflicting incentives, asymmetric information, and distributed decision-making authority, which limits their applicability in practice. Designing coordination mechanisms to resolve existing incentive conflicts is complex and therefore not always considered to be a feasible option. Our work introduces a third path by integrating common real-world conflicts and power differences between organizational units in an overarching decision-support tool. This approach also allows us to capture the implications of different bonus systems in different departments - a common observation in practice.

In addition, our analysis highlights the critical importance of acknowledging future decision flexibility in a medium-term budget planning process. By comparing models based on static RO and AARO, we quantify the value of the real option of being able to adjust production volumes based on observed sales. Our analysis shows that this option value is substantial, which underlines the importance of having processes and information systems in place that ensure rapid availability of latest market information to decision-makers.

Our work opens several opportunities for future research. First, the goal of this paper was to provide an analytic representation of the coordination mechanism used to align functions within the current organizational setup at the case company. We have not aimed at changing this setup but rather at supporting the currently unstructured negotiation process through a formal decision-making approach. Therefore, we have not investigated the optimality of the current coordination scheme and its implications for overall organizational performance. Such an analysis would be highly worthwhile, as would be a discussion of potential coordination schemes for alternative KPIs.

Second, we focused exclusively on the coordination across business functions and assumed that sales and supply chain can each be modeled as a single, global actor. In reality, competing objectives also exist within a function. Extending the proposed model in this respect would be another interesting route for further analysis.

Third, for the sake of focus, we have considered production capacities and inventory investment budget as the only factors that constrain the sales department in minimizing expected lost margins. In practice, strategic considerations imply additional constraints. These include, for example, allocating higher volumes to emerging but currently less profitable markets to foster growth, or introducing minimum supply levels for selected active ingredients to defend market share against substitute products of competitors. Future research could explore how such strategic constraints can be systematically integrated into the model and how they change the results. This would allow for a deeper analysis of the trade-off between short-term profit optimization and long-term strategic goals.

Chapter III

Managing inventory investments in multi-divisional companies: An auction-based approach

with Steffen Klosterhalfen and Moritz Fleischmann

Abstract

Effective inventory management is a critical challenge for corporations and has been the subject of extensive research. In the literature, inventories are typically modeled as a holding cost in the objective function. Our work is based on a research collaboration with a global supplier of agricultural solutions. In contrast to most existing literature, the firm treats inventory not only as a cost component but also as a key performance indicator (KPI) in its own right. To limit total inventory investment while maximizing the firm's expected profit, the management board (MB) establishes a total inventory investment ceiling, which it then allocates across its business units (BUs). However, private information held by self-optimizing BU heads (BUHs) complicates this allocation task.

We propose an auction-based allocation mechanism that converts the inventory investment ceiling into inventory permits, which are then auctioned among BUHs. We analyze the optimal decision-making of BUHs and provide closed-form solutions. We evaluate the approach through two numerical studies based on synthetic data and real-world data from the case company. Results indicate that the proposed auction-based

approach improves expected total profit on average by 4% compared to the firm's current top-down allocation approach. Moreover, we identify non-monotonic relationships among BU characteristics, BU-specific inventory ceilings, and supply quantities, underscoring the importance of leveraging private information through an auction mechanism rather than relying solely on limited central data.

3.1. Introduction

The field of supply chain management provides an extensive body of literature on managing inventories in alignment with firm objectives. In the planning literature, inventories are typically modeled through holding costs within the objective function. However, inventory is not only a cost driver, but also an asset. Financial analysts commonly use the total cash invested in inventories as an independent KPI to evaluate an organization's operating efficiency (Agrawal and Osadchiy 2024, Lai and Xiao 2018). Several authors, including Hendricks and Singhal (2009), have statistically linked high levels of cash tied up in inventories to negative stock market reactions. As a result, management teams consider inventory investment levels as an independent KPI that must be actively managed to meet investor expectations and those of the broader financial market (Bhagwat and Sharma 2007, Lai and Xiao 2018).

Our paper assumes this asset perspective on inventory and addresses the resulting allocation problem. The analysis is based on a research collaboration with a global provider of agricultural solutions. In addition to total profit, the case company is benchmarked by its investors and financial analysts against core competitors based on its inventory efficiency, i.e., the total inventory investment required to enable its operations. As a result, the MB must manage a balanced scorecard (BSC) that includes both total profit and the capital tied up in inventories. Due to limited capacity for supplying products to meet highly seasonal and uncertain demand, the case company must carry substantial levels of inventory, making the management of total inventory investments particularly critical. To implement its BSC and steer its inventories, the MB introduces an inventory investment cap for the business year and seeks to maximize total profit without exceeding this inventory ceiling (IC).

To serve the global market for agricultural solutions, the operations of the case company are structured into BUs. Each BU is independently managed by a BUH and serves its own market segment. To implement its maximum inventory condition in this setting, the company must translate its overall IC into inventory limits for each BU. However, distributed decision-making and private information complicate this disaggregation. Our paper addresses the problem of how to distribute the total inventory investment ceiling across BUs to maximize the firm's total expected profit while respecting BU autonomy.

The task of the MB to set appropriate BU-specific inventory investment ceilings is complicated by the private information held by BUHs on relevant parameters such as profitability levels and the distribution of uncertain demand in their market, which is neither known to the MB nor to other BUHs. As BUs compete for a higher share of the total inventory investment budget, BUHs have an incentive to strategically disclose this information, thereby introducing an agency problem between the MB and the BUHs. Our work focuses on how to deal with this agency issue in the disaggregation of the ceiling. Currently, the MB uses an allocation heuristic that only requires centrally available information and does not rely on BUHs to share private information. While this approach resolves the agency issue, the quality of the resulting BU inventory budgets has increasingly been questioned by both the MB and the BUHs. Concerns have particularly arisen in situations where selected BUs lost a substantial share of uncertain market demand due to the need to constrain product supply to adhere to the inventory investment ceiling, while the total inventory investment of other BUs simultaneously remained well below their IC.

The problem of allocating a central resource is not unique to inventory management, but also arises in the allocation of carbon credits, water rights, computing capacity, ad space, and spectrum management (Milgrom 2004). In those settings, auctions have been suggested to coordinate the allocation of a central resource across decentralized agents under private information and competing incentives. We transfer this idea to the above problem setting and explore the application of an auction mechanism to the problem at hand. Specifically, we suggest splitting the total IC into inventory permits, which can be purchased by BUHs during an auction prior to the start of the business year. By allocating the global IC via an auction instead of a central, top-down allocation based on limited information available to the MB, the firm can align the incentives of BUHs with the corporation's goal of maximizing total profit.

Our contributions in this paper are threefold. First, we provide insights into how a large, multinational organization manages its inventory KPIs beyond a pure holding cost perspective to guide its business decisions. We discuss current company practices and the resulting agency problem caused by private information held by decision-makers with competing incentives.

Second, we propose an auction-based allocation approach to split the total inventory investment ceiling into BU-specific ceilings in an environment of private information and unaligned incentives. We derive optimal, closed-form decision policies for the BUHs and analyze core decision drivers.

Third, we numerically benchmark our proposed allocation scheme against the mechanism currently used at the case company, based on both controlled synthetic data and real company data. We find that using real-world data, the proposed approach can increase expected total profit on average by 4% compared to current company practices.

The remainder of this paper is structured as follows. Section 3.2 provides an overview of the related literature. Section 3.3 formally describes the decision problem. In Section 3.4, we propose an auction-based approach to distribute the IC across BUs and derive the resulting decision-making policies of BUHs, which we then analyze. We assess the performance of the proposed approach compared to current practices at the case company based on synthetic and real-world data in Sections 3.5 and 3.6, respectively. Section 3.7 concludes by summarizing the core findings and outlining directions for future research.

3.2. Literature review

In vast parts of the literature on inventory management, inventories are considered through holding costs in the objective function. This narrow cost-based perspective overlooks the role of inventory as an asset on a company's balance sheet, which is tracked as an independent KPI by both internal and external stakeholders (Agrawal and Osadchiy 2024, Lai and Xiao 2018, Schroeder and Krishnan 1976). Only a limited number of studies have treated inventory as an additional, independent KPI alongside total profit within a firm's BSC (Nahmias and Schmidt 1984, Lee 1994, Bhagwat and Sharma 2007, Agrawal and Osadchiy 2024).

Within this limited body of research, the inventory KPI is typically incorporated into the decision problem either by optimizing return on inventory rather than profit maximization (Schroeder and Krishnan 1976, Li et al. 2008) or by introducing inventory as a separate constraint to limit the maximum inventory level or the total inventory investment (Beckmann 1952, Nahmias and Schmidt 1984, Lee 1994, Bretthauer et al.

2006). Our work focuses on the latter and addresses the real-world problem faced by the case company: how to maximize the firm's total expected profit while limiting the capital invested in inventories in a decentralized decision-making context involving the MB and multiple BUs.

Schroeder and Krishnan (1976) argue that viewing inventories as a cost driver is insufficient and propose to interpret stocks as a financial asset. The authors suggest that firms should maximize profit relative to the required inventory investment and derive economic order quantities under a return-on-inventory maximization objective. Similarly, Li et al. (2008) formulate and solve a production planning problem in which a firm allocates limited funds between inventories and investments in fixed assets to maximize the return on capital employed.

Nahmias and Schmidt (1984) derive the optimal order quantity in a multi-item newsvendor problem. The authors maximize expected profit while ensuring compliance with an inventory constraint related to total available storage space or capital available for inventory investment. To solve the decision problem, the authors provide the Lagrangian and propose several heuristics. Beckmann (1952) proposes an approximate solution procedure using Lagrange multipliers to account for limited storage capacities or capital budgets while simultaneously minimizing total cost in a multi-item lot-sizing problem. Subsequent research by Ziegler (1982) and Ventura and Klein (1988) refines this approach by incorporating appropriate bounds and enhancing the search strategy.

Lee (1994) examines a multi-item economic order quantity problem where order quantities and sales prices are jointly optimized to maximize profit. Beyond the impact of inventory holding cost on total profit, the author introduces inventory constraints on both the maximum available storage space and the total inventory investment budget. The problem is solved using geometric programming. Bretthauer et al. (2006) present a model and solution approach for both continuous and integer-variable versions of production and inventory management problems under multiple resource constraints. In their model formulation, inventories are captured both as holding cost within the objective function and through a constraint on the upper limit of total inventory investment.

We contribute to the aforementioned literature by providing real-world insights into how a multinational agricultural solutions provider manages its inventory levels beyond the standard holding-cost-based approach by implementing a maximum total inventory investment ceiling. A key distinction of our work is that all previously discussed studies assume centralized decision-making by a single decision-maker. In practice, however, decision-making power is often distributed across multiple agents. At the case company, operations are divided into distinct BUs, each led by a dedicated BUH. Our work provides insights into how the case company balances total expected firm profit and the total inventory investment across its BUs through the disaggregation of a global inventory investment ceiling into BU-specific limits.

In our problem context, the IC can be interpreted as a shared scarce resource that must be distributed across BUs. The challenge of allocating limited resources across agents has received substantial research attention across management sciences and has been widely applied to various operations research problems, including production planning, network optimization, and inventory management since the 1950s (Patriksson 2008, Kaihara 2001, Benjaafar et al. 2017).

The effective allocation of resources is often complicated by the distributed structure of most modern corporations, where decision-making power is frequently shared between headquarters and regional managers. In such environments, local managers typically hold private information about the benefits of deploying resources in their markets (Harris et al. 1982). Local managers might therefore be incentivized to withhold or strategically manipulate private information to secure favorable access to global resources that align with their own objectives (Alonso et al. 2008, Rajan and Reichelstein 2004). Similarly, the process of allocating the total IC across the BUs of the case company is complicated by private information held by self-optimizing BUHs. We contribute to the literature on resource allocation (RA) under distributed decision-making and private information by providing insights into a real-life allocation problem faced by the case company. We discuss the stakeholders involved, their incentives, decision-making power, and the allocation mechanism currently implemented at the case company.

To ensure effective resource allocation despite private information, the research field of mechanism design has emerged as a sub-field of game theory that develops incentive schemes to align the actions of rational agents with private information to an overall organizational objective. This approach has been widely applied to various resource allocation problems in operations research (Hurwicz 1973).

Groves and Loeb (1979) examine a problem in which BU managers hold private knowledge regarding the productivity of consuming a shared global resource within their BU. The authors propose a truth-inducing incentive mechanism in which BU managers are compensated based on their contribution to overall profit. Karabuk and Wu (2005) analyze the internal capacity allocation game faced by a major US semiconductor manufacturer, where multiple product managers compete for access to limited production capacity. To overcome the conflicts arising from incentivizing product managers purely based on generated profit, the authors propose a scheme that compensates product managers not only for the profit of their own BU but also through an additional side-payment to ensure that coordination benefits are appropriately shared among product managers and to encourage cooperation. Mallik and Harker (2004) extend the problem to a setting with multiple production facilities. Celikbas et al. (1999) introduce an approach to coordinate the marketing and manufacturing departments of a leading electronics and computer manufacturer by implementing penalties for marketing managers who over-forecast demand and manufacturing managers who under-forecast supply volumes to better align demand forecasts with production capacities.

One common approach in mechanism design is to coordinate self-interested agents by introducing internal markets (Katok and Villa 2022). To this end, a substantial body of literature has emerged on how to design such markets, for example, by leveraging auctions.

Harris et al. (1982) address an intrafirm allocation problem by introducing a transfer pricing scheme in which divisional managers, who hold private information on their own productivity, signal their willingness to pay for a limited global resource. In cases of shortages, the allocation of resources is prioritized based on ranked transfer bids. To ensure truthful bidding, managers are compensated based on a fixed salary minus the cost of the resource allocated to them at the chosen transfer price. Kouvelis and Lariviere (2000) propose internal markets through which business functions can offer and receive outputs. The authors apply their approach to different settings, including a multi-newsvendor problem with a shared capacity constraint and the coordination of a multi-stage production system. Kaihara (2001) presents a market-oriented approach to general supply chain coordination problems based on a bidding system, in which agents can submit supply and demand bids for resources. Through an auction mechanism, a resource price is determined to clear the respective resource market. The author conducts a simulation analysis to confirm that the resulting allocation converges to the Pareto-optimal solution. Ertogral and Wu (2000) solve a multi-level, multi-item capacitated lot-sizing problem in an environment of self-interested decision-makers by resolving scheduling conflicts through conflict pricing, thereby allocating the constrained resource to the highest bidder. Malone (2004) proposes a mechanism to coordinate sales and production plant managers subject to planning constraints by using future contracts to trade the guaranteed availability of products at distinct points in the future. McAdams and Malone (2011) further explore this idea through their collaboration with Intel Corporation. In addition to providing a formal market model, the authors discuss risks and potential solutions for market efficiency issues arising from factors such as collusion among agents and negative externalities. Karabati and Yalçin (2014) propose an auction-based market design to simultaneously solve a production scheduling problem and allocate inventories to buyers.

The above studies primarily address the allocation of scarce physical resources, such as limited raw materials or machine capacities. In our work, we address the allocation of a maximum inventory investment cap, which the case company aims to not exceed. Similarly, BP implemented an internal market to coordinate the disaggregation of their corporate carbon reduction target into individual, BU-specific reduction targets. Initially, BP distributed annual carbon emission permits through a top-down approach. Subsequently, BP opened an electronic trading system that allowed BU managers to buy and sell permits at derived market prices. In 2001, following the initial allocation, the BP's BUHs traded emission permits worth more than $180 million (Malone 2004).

We contribute to the mechanism design literature by examining how a multi-unit auction can be applied to the coordination problem faced by the case company in disaggregating the total IC into BU-specific ICs under conditions of private information and conflicting incentives. To this end, we derive the optimal closed-form decision policies for the BUHs and discuss key drivers of optimal decision-making.

Throughout our work, we assume that agents act rationally and engage in "sincere bidding" by truthfully revealing their valuations during the auction. Therefore, we do not discuss in detail the market design options that could enhance the robustness of allocation outcomes against market efficiency issues. Such issues may arise from strong imbalances in the relative market power of involved bidders, which could provide strategic incentives for large bidders to engage in bid shading (Cramton and Kerr 2002). For further insights into multi-unit auction design across different application environments, we refer to Vickrey (1961), Ausubel (2004) and Milgrom (2004).

3.3. Problem definition and model formulation

In the following section, we provide a detailed description of the decision problem faced by the case company, followed by its formal model representation.

3.3.1. Problem description

Our study is based on a collaboration with an agricultural solutions supplier. The case company provides both chemical and biological crop protection products, as well as seeds and seed treatments, to farmers worldwide. The demand for agricultural solutions is highly seasonal and uncertain, driven by a range of hard-to-predict factors, including global agricultural commodity prices, weather conditions, and local disease pressure.

Due to limited production and processing capacities, the case company pre-produces products to build up sufficient inventories before the selling season. As a result, the company must determine the volumes to be supplied before the realization of uncertain seasonal demand. Any unsold product must be stored until the next selling season. Both pre-production and leftover inventories lead to substantial inventory investments. Pre-production inventories are deterministic, dictated by available capacity and production volumes. In contrast, the cash tied up in leftover inventories depends on actual demand realization and is thus uncertain.

The case company has the option to reduce leftover inventories after the selling season through various measures, such as selling leftover seeds into secondary value chains (e.g., livestock feed) or placing crop protection products into distributor inventories. While helping to free up capital tied up in inventory, these strategies generate additional costs.

The company operates through multiple BUs, each managed by a dedicated BUH who is responsible for its own production assets and performance. All BUHs report to the MB of the company, which retains influence over BU decision-making through target-setting and incentive schemes. The MB also engages with capital markets, including shareholders and creditors.

Pre-production and demand uncertainty lead to significant inventory levels. As highlighted in the introduction, inventory management is critical not only for maximizing profit but also because the total cash invested in inventories is considered a core KPI by the firm's investors and is used to assess the firm's operating efficiency. Financial analysts frequently benchmark the case company against key competitors based on total profit relative to overall inventory investment. Consequently, the MB must balance two KPIs: (i) total corporate profit and (ii) total capital investment in inventories.

However, simultaneously optimizing both KPIs presents a challenge: higher inventory investments ensure adequate product supply to capture uncertain demand, while lower capital deployment increases the risk of lost sales. To manage this trade-off, the MB imposes an annual inventory investment ceiling. Within the limits of this constraint, the MB aims to maximize total expected corporate profit. Due to a highly volatile business environment, the case company focuses its steering efforts predominantly on the current business year.

To operationalize the steering of both KPIs, the MB allocates the total inventory investment ceiling across its BUs by setting individual BU-specific inventory investment ceilings. The MB communicates these ceilings to the BUHs prior to the start of the business year. Subsequently, each BUH is incentivized through an incentive scheme to maximize their own profit while ensuring adherence to the allocated inventory investment budget. Our research focuses on how to optimally allocate the total inventory investment ceiling across BUs to maximize the expected total profit of the firm.

This allocation decision is complicated by two factors:

1. Private information: Key parameters, such as profitability and the distribution of uncertain demand, are privately known only by each BUH, as they require direct customer interaction. Such information is neither known to the MB nor to other BUHs. At the case company, the MB relies on central information systems and industry studies to estimate the value of selected parameters, such as monthly production capacities and expected market demands.

2. Conflicting incentives: Each BUH is compensated based on BU-specific profit. Since the inventory investment ceiling limits product supply and thus profit potential, BUHs have an incentive to strategically disclose private information in an attempt to secure a higher share of the inventory budget.

Given these challenges, the MB currently bypasses direct BU input and instead estimates inventory investment needs using market growth expectations from industry studies and production capacities based on historical production data. Based on these estimates, inventory investment ceilings are allocated proportionally to the pre-production inventory volumes associated with the supply of expected market demands across BUs.

However, the current allocation approach, formalized in Section 3.5.1, has led to dissatisfaction among both the MB and BUHs, particularly due to its inability to account for differences in demand uncertainty and profitability across BUs.

In this paper, we therefore develop an alternative method. Inspired by the success in other domains (Malone 2004), we propose an auction-based mechanism to distribute the inventory investment ceiling across BUs.

3.3.2. Model formulation

We formalize the decision problem as follows. The firm consists of $i = 1, ..., I$ BUs. The goal of the firm is to maximize total expected profit (π) while ensuring that the total inventory investment (l) does not exceed the total inventory investment ceiling (τ). We do not consider the decision of how to set τ itself, as this decision depends on several factors, such as shareholder expectations and market conditions, which are beyond the scope of this study.

In line with steering practices at the case company, we focus our analysis on a single business year and assume that all actors adopt this perspective. Figure 3.1 summarizes the sequence of events. Before the start of the business year, the MB must allocate the total inventory investment ceiling (τ) among the BUs, setting BU-specific inventory investment ceilings (τ_i), such that $\tau = \sum_{i=1}^{I} \tau_i$.

Our work focuses on how to optimally split τ into BU-specific ceilings.

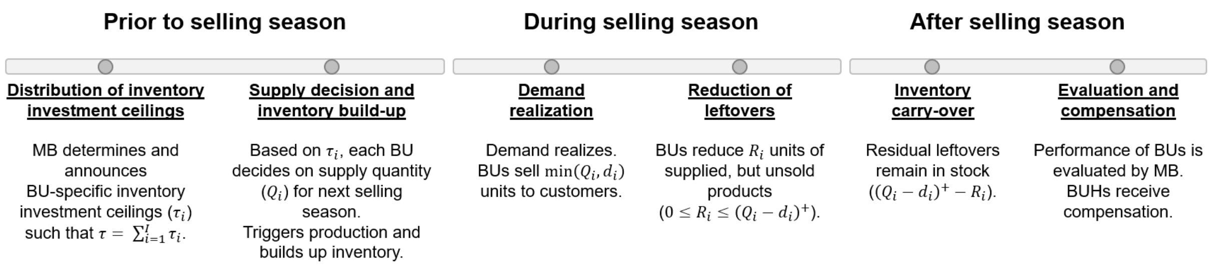

Figure 3.1.: Overview of sequence of events along the business year.

The demand for agricultural solutions is highly seasonal, driven by the annual growing cycle and narrow, product-specific application windows that last only a few weeks each year. We therefore assume that all annual demand of BU i is realized at a single point in time. Consequently, the business year can be divided into t_i^b months before and t_i^a months after the selling season, such that $t_i^b + t_i^a = 12$ months. In addition, we simplify the analysis by aggregating the demand for the individual solutions offered by BU i into a single product representing the total market demand faced by BU i. The total market demand of BU i, denoted as d_i, is uncertain ($d_i \in \mathcal{D}_i$), with a distribution function (DF) of finite support $f_i(x)$ and a corresponding cumulative distribution function (CDF) $F_i(x)$. We assume that market demand is independent across BUs.

Notation	**Description**
Deterministic parameters:	
c_i	Production capacities available at BU i in units per month.
m_i	Per unit profit of BU i.
v_i	Per unit inventory value of BU i.
h_i	Inventory holding cost per unit of inventory value and month of BU i.
k_i	Cost per unit of reduced leftover inventories after realization of demand of BU i.
s_i	Units of starting inventory at beginning of business year of BU i.
t_i^b	Number of months between start of business year and start of selling season of BU i.
t_i^a	Number of months between end of selling season and end of business year of BU i.
τ_i	IC of BU i (transformed from parameter to decision variable for BUHs during auction).
Stochastic parameters:	
d_i	Uncertain annual market demand ($d_i \in \mathcal{D}_i$) faced by BU i with DF $f_i(x)$ and CDF $F_i(x)$.
Decisions variables:	
Q_i	Total product volume supplied by BU i to its market for business year.
R_i	Total volume of reduced leftover inventories by BU i after realization of demand.
π_i	Expected profit of BU i.
l_i	Total inventory investment of BU i.

Table 3.1.: Overview of used notation.

Each BUH aims to maximize the expected profit (π_i) of their BU while ensuring that the total inventory investment (l_i) does not exceed the allocated ceiling (τ_i). In line with the introductory motivation, we therefore treat inventory not only as part of the profit function through holding costs, but also as an asset with a distinct KPI of its own. Once the MB has communicated the allocated inventory investment ceilings, each BUH determines the volume (Q_i) to be supplied to their market for the upcoming selling season.

BUs operate in a continuous production environment, where production is capacity-constrained at c_i units per month. Given an initial stock of s_i units, the BU must ramp up production $\frac{(Q_i - s_i)^+}{c_i}$ months in advance of the selling season. Pre-produced product volumes are kept in BU inventories.

Each unit sold generates a profit of m_i. Any unsold product results in leftover inventory, which can either be carried over to the next selling season or eliminated through dedicated reduction measures. We denote the units of reduced leftover inventory after demand realization as $R_i(d_i)$. Reducing leftover inventory generates an extra cost of k_i per unit. Any remaining unsold product is held in stock until the next selling season, thereby contributing to the total annual inventory investment (l_i) for the remaining t_i^a months of the business year. Figure 3.2 illustrates the relationship between key parameters, decisions, and inventory levels.

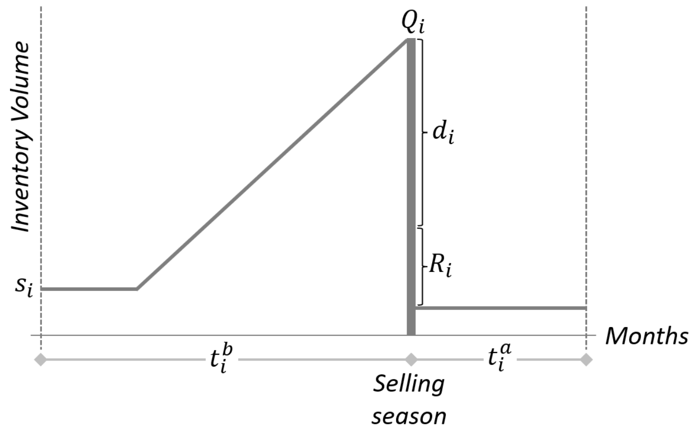

Figure 3.2.: Inventory level along business year.

Each unit of product has a per-unit inventory value of v_i. The total inventory investment (l_i) of BU i consists of two parts: deterministic pre-production inventory held before the selling season and unsold, non-reduced leftover products that remain in stock until the end of the business year. Thus, the total inventory investment is given by:

$$l_i = v_i \cdot (t_i^b \cdot s_i + \frac{(Q_i - s_i)^2}{2 \cdot c_i} + t_i^a \cdot (Q_i - d_i - R_i(d_i))^+). \tag{3.1}$$

If the BU invests capital in inventory, it forgoes returns that could be gained from alternative investments. We denote h_i as the opportunity cost associated with holding one monetary unit worth of inventory for one month. BU i therefore faces a total cost of capital due to inventory investments over the course of the business year of $h_i \cdot l_i$. We assume that the per-unit inventory reduction cost exceeds the capital cost of holding one unit of unsold product from the selling season to the end of the business year, i.e., $k_i > v_i \cdot h_i \cdot t_i^a$.

In the case of centralized information, we can formulate an integrated decision problem in which the MB jointly determines BU-specific inventory investment ceilings ($\vec{\tau} = (\tau_1, \ldots, \tau_I)$) and supply volumes ($\vec{Q} = (Q_1, \ldots, Q_I)$) at the start of the business year, such that total expected profit of the firm is maximized while ensuring that none of the BUs exceed their respective inventory investment ceilings. As discussed in Section 3.3.1, the problem is complicated by information asymmetry. We assume that expected market demands ($E[d_i]$) and monthly production capacities (c_i) are known to the MB, and that all other parameters are private information of the respective BUH, unknown to both the MB and other BUHs. We summarize the notation used in Table 3.1.

Integrated inventory ceiling allocation problem

Under central information availability, the integrated inventory ceiling allocation problem (ICAP) can be formulated as a two-stage stochastic optimization model with recourse:

$$\max_{\vec{Q},\vec{\tau}} \quad E_{\vec{d}}[\pi] = \sum_{i=1}^{I} E_{d_i}[\pi_i(\tau_i, Q_i, d_i)] \tag{3.2}$$

$$\text{s.t.:} \quad \sum_{i=1}^{I} \tau_i \leq \tau \tag{3.3}$$

$$\tau_i \geq 0 \quad \forall i \in \{1, ..., I\} \tag{3.4}$$

$$Q_i \geq s_i \ \forall i \in \{1, ..., I\} \tag{3.5}$$

with

$$\pi_i(\tau_i, Q_i, d_i) = \max_{R_i} \{m_i \cdot min\{Q_i, d_i\} - h_i \cdot l_i - k_i \cdot R_i)\} \tag{3.6}$$

$$\text{s.t.:} \quad l_i = v_i \cdot (t_i^b \cdot s_i + \frac{(Q_i - s_i)^2}{2c_i} + t_i^a \cdot (Q_i - d_i - R_i)^+) \tag{3.7}$$

$$l_i \leq \tau_i \tag{3.8}$$

$$R_i \geq 0 \tag{3.9}$$

$$\tau_i \in \mathbb{R}^+, \, Q_i \in \mathbb{R}^+, \, d_i \in \mathcal{D}_i \tag{3.10}$$

Objective function (3.2) to constraints (3.5) form the first stage, and objective function (3.6) to constraint (3.9) constitute the second stage of the stochastic decision problem with recourse. In the first stage, optimal supply volumes and BU-specific inventory investment ceilings are determined. Objective function (3.2) represents the firm's total expected profit, which is to be maximized and calculated as the sum of expected profit of each BU. Constraint (3.3) ensures that the sum of the BU-specific inventory investment ceilings does not exceed the firm's total inventory investment ceiling. Constraints (3.4) ensure that all BU-specific inventory investment ceilings are non-negative, and constraints (3.5) enforce that each BU must supply at least its initial inventory to its respective market

In the second stage, the optimal volume of leftover stocks R_i to be reduced after the realization of uncertain market demand d_i is determined so that the BU's profit is maximized while ensuring adherence to the inventory investment ceilings. Objective function (3.6) represents the profit of BU i, which consists of realized margins minus inventory holding and inventory reduction costs. Constraint (3.7) specifies the total inventory investment, and constraint (3.8) enforces that it cannot exceed the allocated inventory investment ceiling. Constraint (3.9) ensures that the reduced leftover volume is non-negative.

Note that the model captures the complexities associated with maximizing total expected profit while simultaneously disaggregating the total inventory investment budget into BU-specific inventory investment ceilings, which must not be exceeded under any realization of market demand. Mathematically, the admissible inventory budget is captured by constraints (3.3), (3.4), and (3.8). Omitting these constraints turns the model into a traditional inventory planning problem, in which inventory levels are set based on the trade-off between holding costs and the risk of lost sales. In the absence of an inventory investment constraint, the optimal inventory reduction quantity is zero ($R_i = 0$), since reducing leftover stocks is more expensive than holding inventory until the end of the business year ($k_i > v_i \cdot h_i \cdot t_i^a$).

We can simplify the ICAP by determining the optimal decision policy of R_i in the second stage based on the following proposition.

Proposition 3.1. *Let* $\tau_i \geq v_i \cdot (t_i^b \cdot s_i + \frac{(Q_i - s_i)^2}{2 \cdot c_i})$*, then the optimal leftover reduction decision* $R_i(d_i)$ *after the realization of* d_i *is*

$$R_i(d_i) = (\lambda_i - d_i)^+ \tag{3.11}$$

with

$$\lambda_i \equiv Q_i - \frac{\tau_i - v_i \cdot (t_i^b \cdot s_i + \frac{(Q_i - s_i)^2}{2 \cdot c_i})}{v_i \cdot t_i^a}. \tag{3.12}$$

We provide the proof of Proposition 3.1 in Section B.1 of Appendix B. Note that λ_i represents the lowest demand realization at which all resulting leftovers $(Q_i - \lambda_i)$ can be kept in stock without exceeding the total IC τ_i. λ_i can thus be interpreted as the inventory reduction threshold, i.e., any additional leftover volumes caused by demand realizations below λ_i must be fully reduced to ensure adherence to the IC. Based on Proposition 3.1, we can reformulate the ICAP as a single-stage stochastic problem.

The solution to the integrated problem formulation represents the first-best solution of the ICAP. To solve the ICAP centrally, the MB must possess full knowledge of all relevant problem parameters. However, as motivated in previous sections, relevant parameters are private knowledge of self-optimizing BUHs and we assume that only expected market demands and monthly production capacities are known to the MB. All other parameters are private information of the responsible BUH and are neither known to the MB, nor to other BUHs.

To incorporate information asymmetries and the distributed decision-making structure of the problem summarized in Figure 3.1, we split the problem into I lower-level decision problems faced by each individual BUH and a single upper-level problem faced by the MB.

Decision problem of the BUHs: Optimal supply volume under an exogenous inventory investment ceiling

In the lower-level problem, each BUH decides on their optimal supply and leftover reduction quantities to maximize the expected profit of their respective BU, given a specified inventory investment ceiling. Note that in the lower-level problems, $\vec{\tau}$ is no longer a decision variable, but a parameter set by the MB in the upper-level problem. We formulate the optimal supply response problem (SRP), which is solved by each BUH i for a given inventory ceiling τ_i, as follows:

$$\max_{Q_i} \quad E_{d_i}[\pi_i(\tau_i)] = E_{d_i}[m_i \cdot min\{Q_i, d_i\} - h_i \cdot l_i - k_i \cdot R_i] \tag{3.13}$$

$$\text{s.t.: } Q_i \geq s_i \tag{3.14}$$

$$R_i = (Q_i - \frac{\tau_i - v_i \cdot (t_i^b \cdot s_i + \frac{(Q_i - s_i)^2}{2 \cdot c_i})}{v_i \cdot t_i^a} - d_i)^+ \tag{3.15}$$

$$l_i = v_i \cdot (t_i^b \cdot s_i + \frac{(Q_i - s_i)^2}{2 \cdot c_i} + t_i^a \cdot (Q_i - d_i - R_i)^+) \tag{3.16}$$

$$\tau_i \geq v_i \cdot (t_i^b \cdot s_i + \frac{(Q_i - s_i)^2}{2 \cdot c_i}) \tag{3.17}$$

Objective function (3.13) represents the expected BU-level profit to be maximized. Constraint (3.14) ensures that each BU must at least supply its starting stocks. Equation (3.15) sets the optimal leftover reduction decision based on Proposition (3.1) and equation (3.16) sets the associated total inventory investment. Constraint (3.17) enforces that deterministic pre-production inventories cannot exceed the total inventory investment ceiling to ensure that robust adherence can be ensured under any realization of uncertain demand by reducing leftover inventories.

Proposition 3.2. *Let $v_i \cdot t_i^b \cdot s_i \leq \tau_i$, then the optimal supply quantity of BU i under a given τ_i of the SRP is:*

$$Q_i^* = \begin{cases} min(Q_i^{int}, \overline{Q}_i) & s_i \leq min(Q_i^{int}, \overline{Q}_i) \\ s_i & Q_i^{int} < s_i \leq \overline{Q}_i \end{cases} \tag{3.18}$$

with

$$\overline{Q}_i \equiv \sqrt{2 \cdot c_i(\frac{\tau_i}{v_i} - t_i^b \cdot s_i)} + s_i \tag{3.19}$$

$$Q_i^{int} : m_i + \frac{v_i \cdot h_i \cdot s_i}{c_i} = m_i \cdot F_i(Q_i) + v_i \cdot h_i \cdot (\frac{Q_i - s_i}{c_i} + t_i^a \cdot F_i(Q_i)) + (1 + \frac{Q_i - s_i}{c_i \cdot t_i^a}) \cdot (k_i \cdot v_i \cdot h_i \cdot + t_i^a) \cdot F_i(\lambda_i) \tag{3.20}$$

Note that if $v_i \cdot t_i^b \cdot s_i > \tau_i$, the provided inventory investment ceiling τ_i is not sufficient to cover the deterministic pre-season inventory investment linked to the starting stocks s_i of BU i, rendering the problem infeasible. We provide the proof of Proposition 3.2 in Section B.2 of Appendix B.

By analyzing the optimality conditions, we can derive key insights into the optimal supply decisions of BUHs in response to a given inventory investment ceiling.

Observation III.1. *The optimal supply decision is non-decreasing in the inventory investment ceiling.*

Following its definition, and assuming all else equal, λ_i is decreasing in τ_i. Based on the optimality condition (3.20), the optimal Q_i is non-increasing in λ_i, and therefore non-decreasing in τ_i.

Observation III.2. *For a given inventory investment ceiling, the optimal supply quantity is increasing in margin and production capacity and decreasing in inventory value, inventory holding cost, and inventory reduction cost.*

We can rearrange the optimality condition (3.20) to find that the expected marginal benefit of increasing the supply quantity $(m_i \cdot (1 - F_i(Q_i)) \cdot Q_i)$ equals the marginal increase in expected cost under the optimal Q_i. The marginal cost increase consists of the additional inventory holding cost from pre-production $(v_i \cdot h_i \cdot (\frac{Q_i - s_i}{c_i}))$, the increase in expected inventory holding cost from carrying non-sold and non-reduced stocks until the end of the business year $(v_i \cdot h_i \cdot t_i^a \cdot (F_i(Q_i) - \frac{\partial \lambda_i}{\partial Q_i} \cdot F(\lambda_i)))$, and the marginal cost increase of (partially) reducing leftover inventories $(k_i \cdot \frac{\partial \lambda_i}{\partial Q_i} \cdot F(\lambda_i))$, where $\frac{\partial \lambda_i}{\partial Q_i} = 1 + \frac{Q_i - s_i}{c_i \cdot t_i^a}$. Q_i is thus increasing in m_i and c_i, and decreasing in v_i, h_i, and k_i.

Decision problem of the MB: Split of the overall inventory investment ceiling across BUs

In the upper-level problem, the MB decides how to allocate the overall inventory investment ceiling (τ) into BU-specific ceilings $(\vec{\tau})$, such that the firm's total expected profit is maximized (objective function (3.21)), the total inventory investment ceiling is not exceeded (constraint (3.22)), and all BU-specific inventory ceilings are non-negative (constraints (3.23)). We formulate the inventory ceiling splitting problem (CSP), which is solved by the MB, as follows:

$$\max_{\vec{\tau}} \quad E_{\vec{d}}[\pi] = \sum_{i=1}^{I} E_{d_i}[\pi_i(\tau_i)] \tag{3.21}$$

$$\text{s.t.:} \quad \sum_{i=1}^{I} \tau_i \leq \tau \tag{3.22}$$

$$\tau_i \geq 0 \quad \forall i \in \{1, ..., I\} \tag{3.23}$$

Due to the private information held by each BUH, the MB cannot solve the CSP, as it cannot determine the expected profit resulting from BUHs solving their SRPs in response to a given $\vec{\tau}$. At the case company, the MB relies on an allocation heuristic based on the limited centrally available information, which we present in Subsection 3.5.1.

Note that the coordination problem between the MB and the BUHs arises because the BUHs compete for a higher share of the total inventory investment ceiling to maximize their own BU profits. In the absence of an inventory investment constraint, there would be no need for the MB to intervene, as self-optimizing BUHs would collectively optimize the MB's overall objective function, which is composed of the sum of expected individual BU profits. Without an inventory investment constraint, the inventory level of each BU would be driven by the traditional trade-off between inventory holding cost and the risk of lost sales.

3.4. Auction-based distribution of inventory investment ceiling

To mitigate the risk of suboptimal top-down allocations based on the MB's limited information, we propose eliminating the top-down split of the total investment ceiling into BU-specific targets altogether.

In various domains, auction mechanisms have been successfully applied to allocate limited central resources among self-interested agents with private information. For example, carbon emission auctions are widely used to distribute a fixed cap on carbon emissions across private and public organizations as a means to combat climate change.

During such auctions, bidders purchase carbon emission permits that entitle them to emit a predefined amount of carbon per permit purchased (Cramton and Kerr 2002).

3.4.1. Auction mechanism

We propose replacing the MB's central allocation of inventory investment ceilings with an auction-based mechanism through which BUHs are charged for access to the total inventory investment budget. Instead of assigning investment ceilings to BUs top-down, the total inventory investment ceiling is split into τ inventory investment permits. Each permit entitles its owner to invest one monetary unit in inventory for one month. These permits are auctioned to BUHs through a multi-unit auction, conducted by the MB before the start of the business year. Note that we do not discuss specific auction designs in the following. For an overview of different multi-unit auction designs, see Vickrey (1961), Ausubel (2004), and Milgrom (2004). Instead, we focus on how an auction can be applied to coordinate the problem faced by the case company. Specifically, we assume that BUHs act rationally and engage in "sincere bidding", meaning that they reveal their true valuation of inventory investment permits, and that these permits are therefore allocated to the BUH with the highest willingness to pay under the chosen auction mechanism. While this is a common assumption in the literature, authors such as Vickrey (1961) have shown that sincere bidding might not always be the optimal strategy, particularly in markets with large, dominant bidders. We refer the reader to Ausubel et al. (2014) for a general discussion on the risk of inefficient market outcomes in multi-unit auctions.

In the following, we describe how a multi-unit ascending-clock auction can be applied by the MB to allocate permits across BUHs.

At the beginning of the auction, the MB announces an initial starting price and asks the BUHs to submit bids on the number of permits they intend to purchase at the current per unit price of p. Subsequently, the MB continuously increases the auction price p in discrete steps of ϵ. After each increase, the MB collects bids from each BUH based on the current auction price p. The auction terminates when the total bids no longer exceed τ. At this point, each BUH is allocated the number of permits corresponding to their winning bids, which define its BU-specific inventory investment ceiling (τ_i) for the upcoming business year. The steps of the multi-unit ascending-clock auction are summarized in Algorithm 3.1.

Algorithm 3.1: Auction mechanism: Allocate τ inventory permits across BUs.

```
Result: τ⃗ = (τ_1, ..., τ_I) and p
p ← 0;
Collect new bids based on p from BUs: τ_i(p);
while Σ_{i=1}^{I} τ_i(p) > τ do
    p ← p + ϵ;
    Collect new bids based on p from BUs: τ_i(p);
    τ⃗ ← (τ_1(p), ..., τ_I(p));
Auction terminates with allocation τ⃗ at price p
```

To ensure an efficient allocation of permits and incentive compatibility, BUHs are compensated based on their BU profit minus the permit purchase cost accrued during the auction, i.e., $\pi_i - p \cdot \tau_i$ at the end of the business year. This compensation structure ensures that permits are allocated to BUHs with the highest willingness-to-pay, i.e., those that can generate the highest marginal increase in expected profit from receiving additional inventory investment permits.

By introducing a permit auction, the BUH's objective of maximizing their own expected profit aligns with the MB's objective of allocating inventory permits in a way that maximizes the firm's total expected profit. Thus, the auction mechanism allows the MB to implement the first-best allocation despite decentralized decision-making and the private information held by the BUHs.

Furthermore, the market-clearing price p determined during the auction provides valuable insights for the MB regarding the trade-off between the firm's total expected profit and its maximum inventory investment ceiling. We discuss this trade-off further in Section 3.6.

In the following, we derive and analyze the optimal decision-making policies of each BUH in order to maximize their own expected profit. In Subsection 3.3.2, we first determine the optimal BU response to a given inventory investment ceiling τ_i. In Subsection 3.4.2, we subsequently analyze the optimal decision-making policies when BUHs set their own inventory investment ceilings by acquiring permits through an auction.

3.4.2. Optimal supply decision under an endogenous inventory ceiling

In the following subsection, we determine the optimal bidding strategy of BUHs during the inventory investment permit auction at the beginning of the business year. Each BUH must decide how many permits τ_i to bid for at a given auction price p, as well as the corresponding optimal supply quantity Q_i, to maximize their own expected profit.

We reformulate the decision problem from the previous subsection by converting τ_i from a parameter to a decision variable, and by incorporating the permit purchase cost into the objective function:

$$\begin{aligned} \max_{Q_i,\tau_i} \quad & E[\pi_i] - p \cdot \tau_i \\ & = m_i \cdot E[min\{Q_i, d_i\}] \\ & - h_i \cdot min\{\tau_i, v_i \cdot (t_i^b \cdot s_i + \frac{(Q_i - s_i)^2}{2 \cdot c_i} + t_i^a \cdot (Q_i - d_i)^+\} \\ & - k_i \cdot E[(\lambda_i - d_i)^+] - p \cdot \tau_i \end{aligned} \tag{3.24}$$

$$\text{s.t.: } v_i \cdot (t_i^b \cdot s_i + \frac{(Q_i - s_i)^2}{2 \cdot c_i}) \leq \tau_i \tag{3.25}$$

$$s_i \leq Q_i \tag{3.26}$$

For an auction price of $p = 0$, the model above represents a decision problem without a maximum inventory investment constraint in which the impact of inventory is captured purely through holding cost in the objective function.

Proposition 3.3. *Let*

$$A \equiv m_i + \frac{h_i + p}{c_i} \cdot v_i \cdot s_i \tag{3.27}$$

$$B(x) \equiv p \cdot v_i \cdot t_i^a + (m_i + v_i \cdot h_i \cdot t_i^a) \cdot F_i(x) + \frac{h_i + p}{c_i} \cdot v_i \cdot x \tag{3.28}$$

$$C(x) \equiv (m_i + k_i) \cdot F_i(x) + (p + h_i) \cdot \frac{v_i \cdot x}{c_i} \tag{3.29}$$

Then the optimal decision policies of BUHs during the inventory permit auction are as follows:

Case		***Optimal decisions*** Q_i^*	λ_i^*	τ_i^*
$max\{A, B(s_i)\} \geq B(F_i^{-1}(\frac{p \cdot v_i \cdot t_i^a}{k_i - v_i \cdot h_i \cdot t_i^a}))$				
1	$\frac{m_i - p \cdot v_i \cdot t_i^a}{m_i + v_i \cdot h_i \cdot t_i^a} \geq F_i(s_i)$	$Q_i : A = B(Q_i)$	$F_i^{-1}(\frac{p \cdot v_i \cdot t_i^a}{k_i - v_i \cdot h_i \cdot t_i^a})$	$v_i \cdot (t_i^b \cdot s_i + \frac{(Q_i - s_i)^2}{2 \cdot c_i} + t_i^a \cdot (Q_i - \lambda_i))$
2	$\frac{m_i - p \cdot v_i \cdot t_i^a}{m_i + v_i \cdot h_i \cdot t_i^a} < F_i(s_i)$	s_i	$F_i^{-1}(\frac{p \cdot v_i \cdot t_i^a}{k_i - v_i \cdot h_i \cdot t_i^a})$	$v_i \cdot (t_i^b \cdot s_i + t_i^a \cdot (s_i - \lambda_i))$
$max\{A, B(s_i)\} < B(F_i^{-1}(\frac{p \cdot v_i \cdot t_i^a}{k_i - v_i \cdot h_i \cdot t_i^a}))$				
3	$\frac{m_i}{m_i + k_i} \geq F_i(s_i)$	$Q_i : A = C(Q_i)$	Q_i	$v_i \cdot (t_i^b \cdot s_i + \frac{(Q_i - s_i)^2}{2 \cdot c_i})$
4	$\frac{m_i}{m_i + k_i} < F_i(s_i)$	s_i	s_i	$t_i^b \cdot s_i \cdot v_i$

We provide the proof of Proposition 3.3 in Section B.3 of Appendix B.

Again, we can analyze the optimality conditions of the different cases to gain general insights into the optimal decision-making of BUHs during the inventory investment permit auction. We focus on Cases 1 and 3, in which it is optimal for the BUs to supply volumes beyond their starting inventories and production is therefore larger than zero ($Q_i > s_i$).

If constraint (3.25) is non-binding (Case 1), it is optimal for BUHs to purchase permits beyond those required to cover deterministic pre-production inventories ($\lambda_i < Q_i$), thereby reducing the risk of costly leftover reduction measures needed to ensure adherence to the inventory investment ceiling. The BUH's decision on λ_i can be interpreted as a newsvendor trade-off: if $\lambda_i < d_i$, the BUH has set λ_i too low by purchasing too many permits to cover the risk of additional capital tied up in leftovers. Underage cost corresponds to $p \cdot v_i \cdot t_i^a$ per unit of λ_i below d_i. If $\lambda_i > d_i$, the BUH has set λ_i too high, and it would have been optimal to purchase more permits during the auction to avoid costly reduction measures. Per-unit overage cost corresponds to the leftover reduction cost (k_i) minus the saved permit purchase cost ($p \cdot v_i \cdot t_i^a$), as well as inventory holding cost ($h_i \cdot v_i \cdot t_i^a$). We thus obtain $\lambda_i = F_i^{-1}\left(\frac{p \cdot v_i \cdot t_i^a}{k_i - v_i \cdot h_i \cdot t_i^a}\right)$ based on the critical fractile of the newsvendor trade-off.

We can rearrange the optimality condition of Q_i to $m_i(1 - F_i(Q_i)) = (p + h_i) \cdot v_i \cdot \frac{Q_i - s_i}{c_i} + v_i \cdot t_i^a \cdot (p + h_i \cdot F_i(Q_i))$. The optimal Q_i balances the expected marginal increase in gross profit (left-hand side) with the expected marginal increase in cost (right-hand side).

Observation III.3. *Under Case 1, the optimal supply quantity is increasing in margin and capacity and decreasing in inventory holding cost, inventory value, and the auction price. The optimal supply quantity is independent of the leftover reduction cost.*

Observation III.4. *Under Case 1, the optimal number of purchased permits is increasing in margin and leftover reduction cost and decreasing in holding cost and the auction price.*

Under Case 3, it is optimal for BUHs to only purchase permits to cover the certain pre-production inventories and to ensure adherence to the inventory ceiling by reducing any leftover inventories. We can rearrange the optimality condition to obtain $m_i(1-F_i(Q_i)) = (v_i \cdot h_i + p \cdot v_i) \cdot \frac{Q_i - s_i}{c_i} + k_i \cdot F_i(Q_i)$. Under the optimal Q_i, the marginal benefit of additional expected margin equals the expected marginal cost caused by the holding and permit purchase cost of additional pre-production inventories and the leftover reduction cost if $d_i < Q_i$.

Observation III.5. *Under Case 3 and in line with Case 1, the optimal supply quantity is increasing in margin and capacity and decreasing in holding cost, inventory value, and the auction price. In contrast to Case 1, the optimal supply quantity is decreasing in the leftover reduction cost.*

Observation III.6. *In line with Case 1, the optimal number of purchased permits increases in margin and decreases in holding cost and the auction price. In contrast to Case 1, the optimal number of purchased permits decreases in the leftover reduction cost.*

3.5. Numerical study based on synthetic data

The objective of the following section is to gain insights into the performance of the proposed auction-based mechanism for allocating the overall inventory investment ceiling across BUs, which we denote by AB. To this end, we evaluate its overall performance, analyze the resulting solution structure, and compare it to the current company practice (IP) as well as two additional theoretical benchmarks, VP and EQ.

3.5.1. Study setup and benchmark strategies

We generate a synthetic data set comprising 128 BUs to analyze the performance of the different allocation methods. The $2^7 = 128$ BUs represent the full combination of low and high levels for the seven parameters shown in Table 3.2. Additionally, we assume that demand for all BUs follows a normal distribution with an expected demand and a coefficient of variation as stated in Table 3.2. As discussed previously, our analysis focuses on a single business year. For simplicity, we assume that all demand occurs at the end of month 10, dividing the business year into 10 months before ($t_i^b = 10$) and 2 months after the selling season ($t_i^a = 2$). Furthermore, we assume no starting inventories from the previous season ($s_i = 0$).

Parameter	**Realizations**
$E[d_i]$	$\{50,\ 100\}$
cv_i	$\{0.1,\ 0.4\}$
v_i	$\{5,\ 10\}$
h_i	$\{\frac{0.1}{12},\ \frac{0.3}{12}\}$
m_i	$\{0.25\, v_i,\ 0.75\, v_i\}$
k_i	$\{0.10\, v_i,\ 0.25\, v_i\}$
c_i	$\{\frac{E[d_i]}{0.8\, t_i^b},\ \frac{E[d_i]}{0.4\, t_i^b}\}$

Table 3.2.: Sets of problem parameters used to construct the synthetic data set.

We evaluate the performance of the studied allocation mechanisms across five inventory investment ceiling levels ($ICLs$), which are defined relative to the total inventory investment required to meet the expected market demand across all BUs:

$$\tau^{ICL} = ICL \cdot \sum_{i=1}^{I} v_i \cdot \frac{E[d_i]^2}{2\, c_i}. \tag{3.30}$$

An ICL of 100% corresponds to an inventory investment ceiling equivalent to the total inventory investment required for pre-producing the expected market demand of all BUs. In our analysis, we consider ICLs of 50%, 75%, 100%, 125%, and 150%.

At the case company, the MB currently allocates the total inventory investment ceiling in a top-down manner, setting BU-specific ceilings based on limited central information about expected market volumes ($E[d_i]$) and monthly capacity levels (c_i). Specifically, the MB estimates the pre-production inventory volume required by each BU to meet expected market demand, and distributes the total investment ceiling proportionally across these estimated volumes:

$$\tau_i^{IP} = \frac{\frac{E[d_i]^2}{2\,c_i}}{\sum_{i=1}^{I} \frac{E[d_i]^2}{2\,c_i}} \cdot \tau. \tag{3.31}$$

We refer to the current company practice as IP (inventory-proportional).

Beyond the current company practice, we introduce two additional benchmark strategies. Under volume-proportional allocation (VP), the total inventory investment ceiling is distributed in proportion to the expected market demands across the BUs:

$$\tau_i^{VP} = \frac{E[d_i]}{\sum_{i=1}^{I} E[d_i]} \cdot \tau. \tag{3.32}$$

Under equal distribution (EQ), the total inventory investment ceiling is divided equally among all BUs:

$$\tau_i^{EQ} = \frac{\tau}{I}. \tag{3.33}$$

3.5.2. Performance analysis of allocation strategies

In the following, we present the results of the numerical study, evaluating the performance of the proposed auction-based allocation mechanism (AB) against the current company practice (IP) and the benchmark strategies. Table 3.3 reports the expected total corporate profit under different ICLs for each method. Note that for AB, the total corporate profit is not influenced by auction payments between the BUs and the MB. The percentage values in parentheses indicate the relative difference to the current company strategy (IP). We decompose total expected profit into its three components: the expected gross margin generated from sales to customers, the expected inventory holding cost, and the expected inventory reduction cost required to ensure adherence to the inventory investment ceiling. In addition to presenting the results of the mechanisms under different total inventory investment ceilings, we also provide the results for a pure holding-cost-based model without any constraint on the total inventory investment in the final row of Table 3.3 ($ICL = \infty$). This represents a traditional model setup in which inventory levels are accounted for solely through the holding cost in the objective function, as discussed in Section 3.3.2. The results are equivalent to those of *AB* under an auction price of zero.

	Total profit				Gross margin				Holding cost				Reduction cost			
ICL	***AB***	***IP***	***VP***	***EQ***	***AB***	***IP***	***VP***	***EQ***	***AB***	***IP***	***VP***	***EQ***	***AB***	***IP***	***VP***	***EQ***
Avg.	280.1 (7%)	262.5	264.1 (1%)	258.8 (-1%)	314.4	295.6	296.5	289.8	30.9	29.3	28.9	27.7	3.4	3.8	3.5	3.3
50%	248.6 (13%)	219.2	224.8 (3%)	219.9 (0%)	270.3	239.9	245.8	240.2	17.4	17.7	17.5	17.2	4.4	3.0	3.5	3.2
75%	274.1 (9%)	251.0	255.1 (2%)	248.2 (-1%)	304.9	280.0	283.9	275.5	25.1	25.0	24.6	23.8	5.6	4.0	4.2	3.6
100%	287.1 (6%)	270.8	271.6 (0%)	266.0 (-2%)	323.2	305.5	305.9	298.6	31.8	30.7	30.2	29.0	4.3	4.1	4.0	3.6
125%	293.9 (4%)	282.8	281.5 (0%)	276.7 (-2%)	333.7	322.2	319.2	313.1	37.7	35.1	34.5	32.9	2.2	4.3	3.3	3.4
150%	296.8 (3%)	288.6	287.6 (0%)	283.1 (-2%)	340.0	330.3	327.7	321.7	42.4	38.1	37.5	35.9	0.7	3.6	2.6	2.7
∞	297.7				343.6				45.9				0.0			

Table 3.3.: Expected total corporate profit, gross margin, inventory holding cost, and inventory reduction cost for different methods and *ICLs*.

AB outperforms all benchmark methods across all ICLs. By closing the gap to the first best solution, *AB* improves total expected profit compared to *IP* by 7%, on average. The improvement potential of *AB* is especially strong under low inventory investment ceilings, i.e., 13% expected profit improvement under an ICL of 50%, which reduces to 3% under a high ICL of 150%. These findings emphasize the critical importance of effectively distributing the inventory ceiling, particularly under constrained investment budgets. Simplistic allocation strategies relying solely on limited central information prove inadequate. *AB*, by contrast, incentivizes BUHs to leverage their full private information through bidding strategies in the auction.

Among the benchmark methods, no single strategy consistently outperforms the others. *VP* performs best on average, particularly under low ICLs, whereas *IP* slightly outperforms *VP* at high ICLs. *EQ* performs the worst, especially under high ICLs. As limiting the total inventory investment introduces a constraint to the decision problem, total expected profits are below the expected profits of a pure holding-cost-based model without any inventory investment constraint. As the ICL increases, expected profits converge towards the expected total profit of the holding-cost-based model (297.7). Overall, the results highlight the need for allocation strategies that incorporate BU-specific characteristics beyond centrally available information on expected demand volumes and production capacities.

By analyzing the components of total profit separately, we find the following. *AB* achieves the highest expected gross margin from sales across all ICLs, as the most profitable BUHs purchase a larger share of the permits, particularly under low ICLs. For low and medium ICLs, AB incurs a higher expected reduction cost compared to the benchmark methods, as permits are predominantly used to cover deterministic pre-production inventories to maximize supply and increase gross margin. As a result, a high share of leftovers must be reduced, leading to higher expected reduction costs. If ICLs are high, AB results in a lower expected reduction cost by minimizing instances where underallocated BUs must reduce leftovers while overallocated BUs hold excess permits. Overall, higher ICLs lead to a higher expected inventory holding cost as supply quantities increase, and to a lower expected reduction cost required to ensure adherence to the inventory investment ceiling, which aligns with the results of the pure holding-cost model without an inventory investment constraint. As the per-unit reduction cost exceeds the inventory holding cost of keeping one unit in stock from the selling season to the end of the business year ($k_i > v_i \cdot h_i \cdot t_i^a$), the expected reduction volumes converge to zero under a high ICL.

3.5.3. Solution structure and driver analysis

In the previous section, we analyzed the performance of AB in terms of expected profit. We now examine the solution structure of AB. Specifically, we evaluate the relation between problem parameters and the resulting optimal supply quantity, as well as the optimal number of purchased permits under different ICLs. To this end, we run two sets of linear regression analyses, treating the problem parameters $E[d_i]$, cv_i, v_i, m_i, k_i, c_i, and h_i as independent variables, and the supply quantity Q_i and the number of purchased permits τ_i resulting from AB as dependent variables. We derive the linear regression models for all five ICLs and report the standardized regression coefficients in Table 3.4 and Figure 3.3. The numerical results extend the analytical insights on the directional impact of problem parameters on optimal supply quantities and the number of purchased permits obtained from the analysis of the closed-form solutions in Subsection 3.4.2.

We find that supplied volumes and the number of purchased permits are both increasing in $E[d_i]$ and m_i, and decreasing in h_i across all ICLs. Under low ICLs, BUs with a higher degree of demand uncertainty (cv_i) supply lower volumes and, consequently, purchase fewer permits to cover the associated inventory investment. If the total inventory investment ceiling is tight, more inventory permits are therefore invested in the supply of products serving low-risk markets. Under high ICLs, BUs facing a higher degree of

	Supply quantity (Q_i)							Purchased inventory permits (τ_i)						
ICL	$E[d_i]$	cv_i	v_i	m_i	k_i	c_i	h_i	$E[d_i]$	cv_i	v_i	m_i	k_i	c_i	h_i
50%	0.36	-0.13	-0.32	0.65	-0.03	0.47	-0.09	0.49	-0.16	-0.01	0.81	-0.04	-0.14	-0.09
75%	0.56	-0.09	-0.24	0.49	-0.04	0.36	-0.13	0.67	-0.07	0.10	0.67	0.01	-0.29	-0.16
100%	0.68	-0.01	-0.21	0.40	-0.02	0.27	-0.15	0.73	0.08	0.13	0.57	0.09	-0.35	-0.21
125%	0.74	0.07	-0.18	0.33	0.00	0.22	-0.18	0.75	0.21	0.20	0.46	0.09	-0.37	-0.25
150%	0.76	0.13	-0.15	0.28	0.00	0.19	-0.20	0.76	0.31	0.27	0.37	0.07	-0.37	-0.25

Table 3.4.: Standardized regression coefficients of the linear regression analysis examining the influence of problem parameters on the supply quantity (Q_i) and the number of purchased inventory permits (τ_i) under AB by ICL.

demand uncertainty purchase a disproportionately larger number of permits to serve uncertain but potentially high market demands.

We observe that BUs with a higher per-unit inventory value (v_i) supply lower volumes but are required to purchase more permits to cover the higher inventory investment. BUs with high leftover reduction costs (k_i) supply lower volumes and purchase fewer permits under low ICLs. Under high ICLs, leftover reduction costs have no influence on Q_i but lead to higher permit purchases. The findings are consistent with the analytical insights derived from the closed-form solution in Subsection 3.4.2. Low ICLs lead to high auction prices, and an increasing number of BUs purchase permits only to cover deterministic pre-production inventories (see Case 3 in Subsection 3.4.2). Under higher ICLs and lower auction prices, BUs purchase permits to partially cover the risk of additional inventory investments caused by leftover stocks, i.e., $\lambda_i < Q_i$ (see Case 1 in Subsection 3.4.2). For BUs with a higher k_i, the optimal inventory reduction threshold (λ_i) decreases, leading to a higher number of purchased permits. The optimal number of purchased permits to cover the risk of leftover stocks is thereby set through a separate newsvendor trade-off determining λ_i, independent of the BU's optimal supply volume. Finally, we find that BUs with higher production capacities supply higher volumes to their markets while requiring fewer inventory permits due to lower pre-production inventories.

Based on the standardized regression coefficients, we can gain insights not only into the directional effect of parameter values on supplied volumes and purchased permits, but also assess their relative importance for different ICLs. Under low ICLs, supply volumes of different BUs are primarily differentiated by the profitability and production capacity of the BU. As a result, BUs with high profitability purchase a disproportionately large share of the overall inventory investment ceiling, and the available inventory investment is primarily allocated to BUs with the highest profitability. Note that while BUs with high production capacity supply higher volumes to their markets, they purchase fewer inventory permits because higher production capacities lead to lower pre-production inventories. Under high ICLs, supply volumes are mainly differentiated by expected market size ($E[d_i]$). As overall supply volumes increase with higher ICLs, inventory holding costs become more important since pre-production inventories increase quadratically with Q_i. As a result, BUs with larger expected demand and lower production capacities purchase more permits.

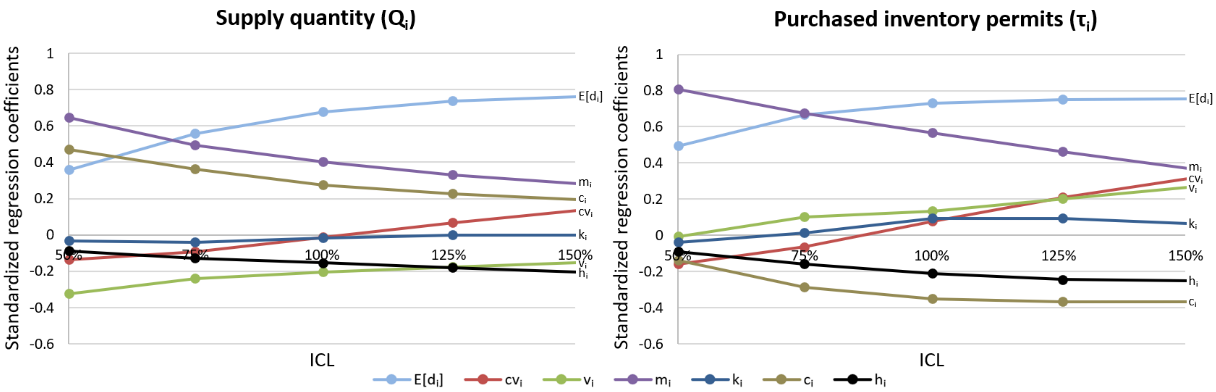

Figure 3.3.: Visualization of regression results presented in Table 3.4.

The analysis reveals several complex relationships between the problem parameters, the optimal supply quantity, the optimal inventory investment ceiling by BU, and the total inventory investment budget. The findings highlight the importance of a sophisticated auction-based allocation mechanism compared to simple benchmarks.

3.5.4. Expected inventory investment levels

When setting the supply quantity and, under AB, determining the number of permits to purchase, each BUH must decide how many permits to reserve to cover the additional capital tied up in potential leftover inventories and thereby avoid the need for costly leftover reduction measures while ensuring compliance with the inventory investment ceiling. This decision relies on a newsvendor trade-off in setting the leftover reduction threshold λ_i to balance overage cost (post-season leftover reductions) and underage cost (excess permits), as discussed in Subsections 3.3.2 and 3.4.2. Consequently, the actual total inventory investment will be below the inventory investment ceiling whenever market demand d_i exceeds the leftover reduction threshold λ_i. Table 3.5 reports the gap between the expected total inventory investment across all BUs ($E[l]$) and the total inventory investment ceiling (τ) for different allocation methods across varying ICLs.

Exp. inv. ($E[l]$) vs. inv. invest ceiling (τ)				
ICL	***AB***	***IP***	***VP***	***EQ***
Avg.	**-6.3%**	**-13.4%**	**-14.7%**	**-17.6%**
50%	0.0%	-1.7%	-2.5%	-4.2%
75%	-1.7%	-6.3%	-7.6%	-10.6%
100%	-5.5%	-12.8%	-14.4%	-17.7%
125%	-10.0%	-19.8%	-21.2%	-24.8%
150%	-14.5%	-26.4%	-27.7%	-30.8%

Table 3.5.: Gap between expected total inventory levels versus total inventory investment ceiling resulting from different mechanisms under different ICLs.

For all studied strategies, expected inventory investments fall below the inventory investment ceiling. Among the studied methods, *AB* consistently results in the lowest gap between the expected total inventory investment and the ceiling. For an ICL of 50%, all permits are used exclusively to cover pre-production inventory, and all leftover stock is fully reduced, leading to a gap of zero. For medium and high ICLs, all benchmark methods exhibit substantially larger gaps than *AB*. *EQ* generates the largest expected undershoot because it fails to account for relevant parameters, causing the strongest simultaneous over- and under-allocation of permits. While gaps between the expected inventory investment and the ceiling do not necessarily translate into lower total profit, they present a challenge for the MB as they constrain its ability to precisely manage the total inventory investment through the introduction of strict inventory investment ceilings.

3.5.5. Inventory efficiency

As outlined in the introduction and problem description, a key rationale behind the MB's implementation of inventory investment ceilings is its strategic view of inventory not merely as a cost factor within the profit function, but as a standalone asset. This perspective drives the MB's efforts to optimize inventory efficiency across its BUs, which is defined as the ratio of generated profit to the capital tied up in inventory. To assess differences in the inventory efficiency resulting from the studied methods, we derive the inventory investment allocations resulting from each method under different ICLs (50% to 150%). For each ICL and method, we derive the expected total corporate profit and the resulting expected total inventory investment. The results shown in Figure 3.4 provide insights into how the allocation mechanism influences the overall inventory efficiency of the corporation. The results confirm that AB outperforms all benchmark methods in inventory efficiency. The advantage is particularly pronounced under low expected inventory investments resulting from tight ICLs. Among the benchmark methods, VP demonstrates superior efficiency relative to the other benchmarks, while the current company practice (IP) performs the worst at low ICLs.

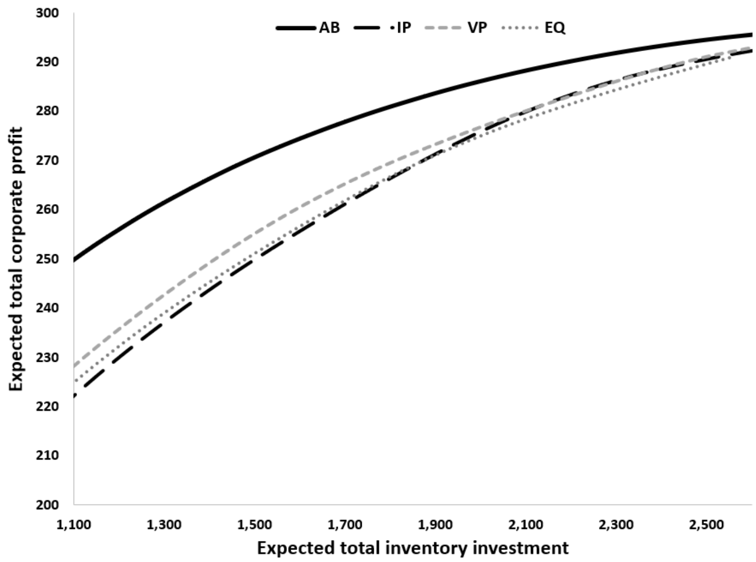

Figure 3.4.: Expected total profit compared to expected total inventory investment by allocation strategy.

3.6. Application to real-world data of the case company

Building on the insights from the synthetic data study, we now assess the improvement potential of the proposed allocation mechanism (AB) in a real-world setting using data from the case company.

To this end, we use real-world data from the case company's 20 BUs. Each BU is managed by a dedicated leadership team and operates its own production assets. We set BU-specific production capacities, profitability levels, and inventory values based on historical averages. Demand uncertainty is assumed to follow a normal distribution, with a BU-specific coefficient of variation derived by comparing historical demand forecasts with actual market demands from past business years. Since the historical per-unit cost of reducing leftover inventories is not readily available in the company's reporting system, we rely on expert judgment obtained through interviews. To protect confidentiality, we anonymize and normalize all input data and results.

3.6.1. Performance analysis of allocation strategies

Table 3.6 summarizes the expected total profit, gross margin, inventory holding cost, and inventory reduction cost for each allocation method under different ICLs, along with the results of a pure holding cost-based model without any constraint on the total inventory investment ($ICL = \infty$), using real-world data.

AB improves the total expected profit compared to the current company practice (IP) by 4%, on average, and outperforms all other benchmarks. However, the relative improvement potential of AB over IP is lower under the real-world data compared to the synthetic data study, particularly under low ICLs. The key reason for this difference lies in the degree of parameter differentiation. The synthetic data set was constructed as a full combination of the presented parameter sets. In contrast to other benchmark strategies, AB can tailor BU-specific investment targets to the high degree of parameter differentiation. For the real-world data, the degree of parameter differentiation across BUs is relatively lower, particularly in terms of profitability, which reduces the relative advantage of AB over the benchmarks.

Nonetheless, a 4% profit improvement based solely on how the global inventory investment ceiling is split across BUs is regarded as significant by the case company.

In contrast to the synthetic study results, IP outperforms all other benchmark methods across all ICLs in the real-world setting. Notably, EQ performs significantly worse than in the synthetic data study, where it was on par with IP and VP for low and medium ICLs. This decline in performance is driven by the greater variation in BU sizes in the real-world data, which makes an equal split of inventory investment ceilings especially insufficient. The findings on the expected gross margin, reduction cost, and holding cost for different strategies and ICLs, as well as the pure holding cost-based model, align closely with those from the synthetic data study.

	Total profit				Gross margin				Holding cost				Reduction cost			
ICL	***AB***	***IP***	***VP***	***EQ***	***AB***	***IP***	***VP***	***EQ***	***AB***	***IP***	***VP***	***EQ***	***AB***	***IP***	***VP***	***EQ***
Avg.	228.0 (4%)	219.3	215.9 (-2%)	199.9 (-9%)	256.6	246.4	241.5	222.5	25.6	24.4	23.1	20.2	3.0	2.6	2.6	2.4
50%	195.6 (5%)	185.6	181.1 (-2%)	168.0 (-9%)	212.4	201.8	196.2	182.3	13.8	14.2	13.3	12.6	2.9	2.0	1.8	1.7
75%	222.2 (5%)	211.9	207.7 (-2%)	190.4 (-10%)	247.0	234.5	228.8	209.8	20.6	20.2	18.9	17.3	4.2	2.4	2.2	2.0
100%	235.6 (4%)	227.1	223.6 (-2%)	204.8 (-10%)	266.3	256.0	250.8	228.8	26.6	25.6	24.1	21.2	4.1	3.3	3.1	2.8
125%	241.9 (3%)	234.2	231.4 (-1%)	214.3 (-8%)	275.8	267.2	262.9	240.9	31.4	29.8	28.2	23.9	2.6	3.3	3.3	2.8
150%	244.8 (3%)	237.9	235.6 (-1%)	221.8 (-7%)	281.6	272.3	268.9	250.7	35.7	32.2	30.9	26.2	1.0	2.2	2.5	2.7
∞	245.9				285.3				39.4				0.0			

Table 3.6.: Expected total corporate profit, gross margin, inventory holding cost, and inventory reduction cost for different methods and $ICLs$.

3.6.2. Expected auction price for different ICLs

Through the auction, we determine a per-unit price at which BUHs can purchase inventory investment permits, ensuring that the total bids match the total inventory investment ceiling. Beyond aligning BUH incentives with the MB's objective of maximizing the firm's expected profit, the auction price provides valuable insights for the MB. The auction price can be interpreted as the shadow price of the inventory investment ceiling constraint. Figure 3.5 shows the numerically obtained auction price across different ICLs. For example, an auction price of 0.02 obtained when issuing an amount of permits that corresponds to an ICL of 100% implies that total expected profit would increase marginally by 0.02 for each additional permit issued. The results highlight the trade-off between the total inventory investment ceiling and the total expected profit, providing the MB with a quantitative basis for evaluating adjustments to inventory investment limits.

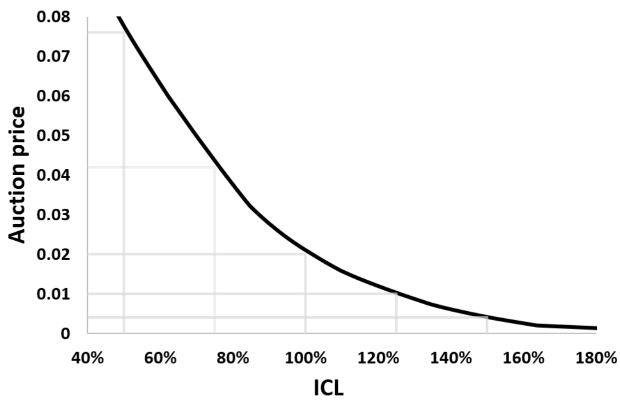

Figure 3.5.: Auction price (p) resulting from issuing different ICLs as inventory investment permits to BUs.

3.7. Conclusion

The majority of inventory models in the literature capture inventory exclusively via holding costs in the objective function. This narrow cost-based perspective overlooks the role of inventory as an asset on a company's balance sheet. In practice, the capital invested in inventories is commonly tracked as an independent KPI, serving as a critical metric for financial analysts and market participants when evaluating a corporation's operational efficiency (Agrawal and Osadchiy 2024, Lai and Xiao 2018). Our work addresses the non-trivial planning problem that arises from treating the total inventory investment as a separate KPI in addition to total profits. It is grounded in a collaboration with a global supplier of agricultural solutions. The case company is benchmarked by its investors based on total profit and total inventory investment against its core competitors. To manage both profitability and inventory investment simultaneously, the MB of the case company imposes a total inventory investment ceiling, which is then allocated across BUs as BU-specific inventory investment ceilings. However, this allocation is complicated by the presence of private information held by self-optimizing BUHs. Our work focuses on how to efficiently convert a total inventory investment ceiling into BU-specific ceilings in this context. To this end, we propose translating the overall inventory investment ceiling into inventory investment permits, which can be acquired by BUs during an auction. By selling inventory permits to the highest bidder, the MB increases the cost associated with binding cash in inventories, thereby ensuring that the right to carry stocks is allocated to the BUs with the highest willingness-to-pay. We formalize the decision problem faced by BUHs during the auction and derive optimal decision policies in closed-form solutions.

Through two numerical studies based on synthetic and real-world data, we evaluate the performance of the proposed auction-based mechanism against the current top-down allocation approach of the case company and additional benchmark strategies.

We find that for real-world data, applying the proposed approach can improve expected total profit on average by 4% compared to current company practices. Moreover, we observe that the profit-optimal allocation of the total inventory investment ceiling across BUs depends on several BU characteristics, many of which are private information held by the BUHs. The relative importance of these characteristics varies with the size of the total inventory investment ceiling, revealing several non-monotonic relationships between BU characteristics, the optimal inventory investment ceiling, and the resulting supply quantity of the BUs. These findings underscore the value of leveraging private information held by BUHs through an auction mechanism, as opposed to allocating inventory ceilings top-down based on limited information available to the MB.

Our study opens several opportunities for further research. First, our work focuses on a myopic strategy in which decision-makers optimize performance within a single business year. However, year-end inventory levels determine starting inventories for the following year. Future research could explore optimal decision-making in a multi-period setting. Second, the goal of this work is to support the current business steering mechanism of the case company, in which the total inventory investment ceiling is distributed into individual BU-specific investment ceilings. Our numerical study shows that this leads to expected inventory investments below the investment ceiling, particularly creating situations in which some BUs engage in costly post-season inventory reductions to ensure adherence to the BU-specific ceiling, while other BUs simultaneously undershoot their inventory investment ceiling due to high market demands. Future research should focus on designing an incentive-compatible pooling mechanism that allows BUs to share unused permits, thereby reducing excessive inventory reductions and further improving total profit. Third, while this study focuses on the two core KPIs of the case company (total profit and total inventory investment), corporate performance is typically assessed using a broader set of financial and non-financial KPIs. Managing these KPIs within a decentralized decision-making framework introduces additional complexities. Extending the analysis to explore how to simultaneously steer a wider range of KPIs would be a highly valuable research direction.

Chapter IV

Allocating inventories in distributed organizations: A hybrid approach applied in the agrochemical industry

with Pardis Sahraei and Moritz Fleischmann

Abstract

We study a multi-resource allocation problem in a distributed organization where a central decision maker (headquarters) lacks direct market demand information and therefore relies on orders from local product managers (PMs) with market access. This information asymmetry, combined with conflicting incentives among self-interested agents, encourages strategic behavior such as order inflation, underscoring the need for an effective allocation mechanism. We propose a novel *base-and-refine* (BR) mechanism, which combines an initial top-down allocation with a second step that allows PMs to swap portions of their allocations through transfer payments based on transfer prices set by the headquarters. Our numerical analysis demonstrates that BR outperforms both, theoretical and practical benchmarks, thereby significantly narrowing the gap to the first-best solution achievable under full information. Furthermore, we show that the two-step design, integrating a top-down allocation with transfer-price-based adjustments, outperforms either step in isolation and is substantially more robust to misspecified transfer prices than a pure transfer-pricing approach.

4.1. Introduction

In today's uncertain and volatile global supply chain environment, business success is tied to the ability of organizations to manage limited resources, such as production and distribution capacities, raw materials, and inventories, through effective allocation decisions. Over the past few decades, a substantial body of literature has addressed a wide range of allocation problems (Luss and Gupta 1975, Curry 1990, Cachon and Lariviere 1999b). Most existing literature proposes integrated, monolithic optimization models to centrally determine optimal resource allocations to be implemented by a single decision-maker (Bretthauer and Shetty 1995, Lau and Lau 1996, Barahona et al. 2005, Claro and De Sousa 2012). Such models typically assume, at least implicitly, a singular, central decision-maker optimizing the single objective function of the organization. However, monolithic model formulations ignore the complexities arising from information asymmetries caused by the structure of large-scale organizations in which decision-making power is distributed across multiple agents. In such an environment, local decision-makers typically possess information that is inaccessible to others. Depending on the incentive structure in place, self-optimizing decision-makers might withhold or distort private information relevant to the allocation problem, causing an adverse selection problem. This study addresses the challenge of allocating central resources among distributed, self-optimizing agents holding private information.

4.1.1. Case description

Our work is based on a collaboration with a global agrochemical corporation that supplies crop protection products (CPPs) to farmers to protect crops against weeds, insects, and fungi (Sparks and Lorsbach 2017). CPPs are produced in a two-step production process. At the case company, both production steps and all intermediate material flows are centrally managed by the headquarters (HQ). In the first production step, active ingredients (AIs) are synthesized in large, capital-intensive facilities, requiring high utilization rates for cost-efficiency. To meet the highly seasonal demand for CPPs, which lasts only a few weeks per year, the case company starts pre-production of AIs up to a year in advance (Shah 2004). Since expanding production capacity requires long-term investments (Fritz and Hausen 2009), available AI volumes are fixed in the short- and medium-term. Close to the selling season, the HQ triggers the second production step, referred to as formulation, in which one or multiple AIs are blended into a final product. At this stage, the HQ faces the allocation problem of deciding which final products to manufacture from the limited AI volumes to maximize the firm's expected profits.

This allocation decision is complicated by high uncertainty in CPP demand, which is influenced by several hard-to-forecast factors, including disease pressure and local farming conditions (Fritz and Hausen 2009). Because evaluating these local drivers centrally is difficult, the HQ relies on distributed product managers (PMs). Each PM is responsible for forecasting demand within their assigned sales territory, placing orders for the required product volumes with the HQ before the selling season, and subsequently fulfilling customer demand. If submitted orders cannot be supplied based on the available AI volumes, the HQ must centrally decide how to allocate AI quantities across final products and PMs.

The allocation problem of the case company is further complicated by information asymmetry and conflicting incentives. Knowledge of local market conditions requires access to local customers. As neither the HQ nor other PMs directly interact with customers of a given PM, the PM is the only one able to generate a probabilistic demand forecast for the following selling season. To obtain some information on the range of plausible demands without direct access to customers, the HQ develops worst- and best-case demand scenarios per product based on centrally available market intelligence studies on the range of expected crop areas and market shares. Other information such as profit margins, warehousing costs, and the bill of material of final products is accessible through a shared controlling system and therefore available to both the responsible PM and the HQ.

To reward sales performance, the PMs of the case company are compensated in proportion to local profits, calculated as gross profits generated from sales to customers minus additional local operational expenses, primarily handling and storage costs for unsold products after the end of the selling season. Based on the compensation scheme, PMs are incentivised to strategically place orders in an attempt to maximize their own local, rather than global, profits, introducing an adverse selection problem between the PMs and the HQ. In the past, prior to the introduction of a formal allocation process, the case company faced significant challenges with PMs strategically inflating their orders to secure a higher share of limited AI volumes, especially when a shortage was feared. This led to mistrust within the organization and a tedious, unstructured, and time-consuming allocation effort, often spanning multiple weeks of iterative negotiations (Harris et al. 1982).

The HQ therefore introduced a formal, two-step lexicographic allocation mechanism, which it announces to all PMs prior to order submission. Once all orders are collected, the HQ verifies whether they can be fulfilled with the available AI volumes. If this is not the case, it initiates the allocation process. In the first step, PMs receive a volume confirmation corresponding to the lower bound of uncertain demand, based on the worst-case scenario estimated by the HQ using centrally available market intelligence. In the second step, volumes are increased to the submitted order until AIs are fully allocated or all orders are met. In both steps, PMs are prioritized based on their relative profitability. We present the approach in more detail in Section 4.4.2. By establishing this mechanism, the case company has streamlined volume allocations and eliminated the incentive for order inflation. However, questions arose regarding the potential negative impact of the allocation method on overall performance. These concerns were particularly voiced by PMs who experienced significant lost sales while observing other, often only marginally more profitable PMs, carrying large volumes of unsold products into the next selling season.

This situation highlights the fundamental dilemma faced by the HQ when designing an allocation mechanism. If allocations are based solely on centrally available information, the HQ does not need to rely on PMs to truthfully share private data. However, this approach risks ignoring critical allocation criteria, most notably demand uncertainty, which is only known locally. As a result, the allocation may be robust against strategic manipulation but suffer from poor overall performance. Conversely, if the HQ seeks to incorporate relevant but decentralized information into the allocation process, it must depend on PMs to truthfully disclose their private knowledge. This introduces the risk of basing decisions on strategically distorted inputs, which can also lead to suboptimal performance outcomes.

The goal of this study is to navigate the above dilemma and enhance the current allocation process of the case company to improve total expected profits in the described context of private information and conflicting incentives. To address this challenge, we propose a hybrid allocation mechanism, referred to in the following as *base-and-refine* (BR), that combines centralized decision-making with decentralized adjustments, enabling the HQ to leverage local information while mitigating strategic misreporting. BR

consists of two allocation steps. In the first allocation step, the HQ determines a top-down base allocation, based on centrally available information. After announcing the base allocation, PMs have the option to swap initially allocated volumes in exchange for a centrally defined transfer payment in the second step. Through the second step, PMs are incentivized to jointly correct the initial top-down allocation based on bottom-up adjustments by leveraging their private information to improve the resulting expected total profit of the firm.

To ensure implementability, we aligned with the case company that any developed allocation mechanism must fulfill the following three properties:

1. Implementable: The allocation mechanism must be implementable under the presented asymmetric information conditions, particularly considering the limited knowledge of the HQ regarding the distribution of uncertain local market demands.

2. Feasible: Any allocation mechanism must fully eliminate shortages for each AI, ensuring that the confirmed product volumes can be produced within the limits of available AI supply.

3. Strategy-Proof: An allocation mechanism is strategy-proof if it is incentive compatible, i.e., each PM's dominant strategy is to order his self-optimizing quantity (Satterthwaite 1979). In this context, strategy-proofness does not require PMs to maximize overall company performance but only to truthfully reveal their self-interested preferences.

4.1.2. Main contributions

Our work offers three key contributions to the existing literature and practice.

First, we provide insights into a real-world allocation problem within the agrochemical industry. We analyze the roles and incentives of the actors involved and present the current allocation heuristics employed by the case company. In addition, we formalize the presented allocation problem through a monolithic, first-best model formulation and critically assess its limited applicability in the studied environment. Based on the discussed problem setting, we point out that many of the approaches proposed in the literature for allocating resources under private information suffer from severe practical limitations. Easy-to-implement mechanisms such as rationing policies tend to be overly simplistic, risking significantly suboptimal allocation outcomes. More sophisticated mechanisms such as internal markets, while theoretically sound, are often too complex and costly to implement in many real-world environments. We argue that this tension between overly simplistic and theoretically sound yet hard-to-implement approaches limits the practical relevance of existing research. This conflict is further amplified in settings where multiple scarce resources must be allocated in parallel, as in our case with multiple AIs, a complexity rarely addressed by simple allocation mechanisms. This underscores the need for solutions that balance ease of implementation and performance.

Second, we propose a novel two-step hybrid allocation mechanism that integrates an initial top-down allocation with bottom-up refinements based on decentralized information. In the first step, the HQ allocates quantities using centrally available data. In the second step, PMs adjust these allocations through a trading process, exchanging part of their initial allocations for a transfer payment set by the HQ. The proposed mechanism incentivizes PMs to truthfully leverage their private information to improve the initial top-down allocation. At the same time, our method is easier to implement than other methods from the literature which, for example, require a complex combinatorial auction to determine allocations. Our hybrid approach addresses the previously discussed dilemma by balancing simplicity with the need to incorporate relevant but distributed information in allocation decisions and is applicable in complex settings where multiple resources must be allocated in parallel.

Third, we test our allocation mechanism numerically against current company benchmarks and other common heuristics. Based on a controlled synthetic data set, we identify key performance drivers and discuss the robustness of the obtained results. We then evaluate the performance of the proposed mechanism based on a real-world data set from the case company.

The remainder of this paper is structured as follows. In Section 4.2, we provide an overview of the relevant literature. In Section 4.3, we formalize the allocation problem and the resulting decision problems of the PMs and the HQ. We provide the monolithic model representation of the allocation problem in Section 4.4.1 to derive the first-best solution and present the heuristic currently employed by the case company in Section 4.4.2. We describe the mechanism proposed by this work in Section 4.4.3. We finally assess the performance of our method compared to current company practice and additional benchmarks based on a synthetic data study in Section 4.5 and real-world data of the case company in Section 4.6. We summarize our work and discuss limitations as well as opportunities for future research in Section 4.7.

4.2. Literature review

Resource allocation has been widely studied across multiple domains such as portfolio management (Cooper et al. 2000), revenue management in hospitality and aviation (Aydin and Birbil 2018, Curry 1990), and inventory management (Hadley and Within 1963, Agrawal and Cohen 2001). In symmetric information environments, integrated optimization models effectively address allocation problems. In asymmetric information settings, potential incentive conflicts can lead to adverse selection, which complicates effective resource allocation and requires appropriate allocation mechanisms (Mishra et al. 2007).

This study contributes to the resource allocation literature by proposing a novel mechanism that combines top-down allocations based on limited information with bottom-up adjustments incentivized by transfer payments. To determine transfer prices based on limited central information, our method leverages the model structure of the capacitated newsvendor problem (NVP). The following subsections review the relevant literature on optimization-based resource allocation, rationing policies, and mechanism design.

4.2.1. Optimization-based resource allocation

Optimization-based techniques for resource allocation and capacity planning have been extensively studied in both deterministic and stochastic contexts. For instance, Bretthauer and Shetty (1995) introduce a general non-linear deterministic resource allocation model that addresses a single capacity constraint, with applications in manufacturing capacity planning and capital budgeting, and solve it using a branch-and-bound algorithm. In stochastic settings, models incorporate demand uncertainties through scenario-based methods, such as Barahona et al. (2005) for semiconductor planning, Claro and De Sousa (2012) for pharmaceutical capacity management under regulatory uncertainties, and studies on capacity allocation in the TFT-LCD industry by Lin et al. (2011) and Chen and Lu (2012), which employ stochastic mixed-integer programming models.

Within the inventory management domain, the NVP has emerged as a classical stochastic model for determining the optimal inventory level of a single product prior to the realization of uncertain demand by balancing expected overage and underage costs (Birge and Louveaux 2011, Maggioni et al. 2019). In many real-world problems, such as the one discussed in this study, firms must optimize the inventory levels of multiple products in parallel while accounting for additional resource constraints. Authors have therefore extended the traditional NVP to more complex scenarios. Lau and Lau (1996) propose a solution procedure for the multi-product NVP with a single resource constraint and general demand distributions. Erlebacher (2000) derive a closed-form solution for a multi-item NVP subject to a single capacity constraint. Zhang et al. (2009) solve a multi-item NVP with multiple non-negativity constraints and a single budget constraint. Building on this work, Zhang (2012) extends the multi-product NVP to multiple capacity constraints and introduces a multi-tier binary search method grounded in Karush-Kuhn-Tucker (KKT) conditions. In our study, we model the allocation problem faced by the case company as a multi-capacitated NVP, incorporating underage and overage costs for each final product while respecting availability constraints for multiple AIs.

While such monolithic optimization models are effective in symmetric information environments, they fail in asymmetric settings where key input parameters remain private to specific agents (Karabati and Yalçin 2014, Kim et al. 2021). In modern, large-scale organizations, distributed decision-making power often results in relevant information not being equally accessible to all agents (Karabuk and Wu 2005). For the problem at hand, local PMs are the only ones with detailed knowledge of the likelihood of specific demand realizations. As PMs are incentivized by their own profits, the allocation problem is complicated by their strategic use of private information to secure a larger share of limited AI volumes. The HQ can address the challenge of private information held by distributed PMs either by applying simple allocation heuristics, such as rationing policies, that rely solely on centrally available information, or by designing allocation mechanisms that incentivize PMs to truthfully disclose their private information. In the following subsections, we review the relevant literature on rationing policies and mechanism design.

4.2.2. Allocation heuristics

Rationing policies have been extensively studied and applied in practice, especially in industries characterized by long lead times, uncertain demand, and inflexible capacity, such as electronics, automotive, and pharmaceuticals (Li et al. 2017, Hofstra and Spiliotopoulou 2022). Rationing policies typically require minimal centrally available information, making them particularly suitable for environments in which distributed agents possess private information (Cachon and Lariviere 1999b). Common rationing policies include linear, proportional, uniform, lexicographic, and turn-and-earn allocation rules. Each mechanism uses different inputs to determine allocation outcomes and creates distinct incentives for strategic behavior among agents. In the following subsection, we present and compare the most commonly used rationing policies.

Under linear allocation, a fixed quantity is deducted uniformly from all orders submitted by agents. Proportional allocation allocates resources proportionally to submitted order volumes. Both methods naturally incentivize order inflation as agents aim to secure a higher allocation (Lee et al. 1997, Cachon and Lariviere 1999b). Turn-and-earn policies base future allocations on actual sales from previous periods and have found widespread application in the supply chain of fast-moving consumer goods and in the automotive industry (Lu and Lariviere 2012, Cachon and Lariviere 1999a). As turn-and-earn bases allocations on past sales, it does not incentivize order inflation. Under uniform allocation, the available capacity is equally divided among agents, with each agent receiving the minimum of their original order quantity and an equal share, irrespective of other characteristics such as profitability (Cachon and Lariviere 1999c). Lexicographic allocation policies fulfill orders based on a predetermined priority list, which can be constructed using criteria such as relative profitability or strategic importance of the agents (Cachon and Lariviere 1999b, Chen et al. 2013). Both methods are robust against gaming, as participants cannot increase their allocation through strategic behavior such as order inflation.

While rationing policies are easy to implement, their appropriateness for resolving a given allocation problem must be carefully assessed. For the problem studied in this paper, linear and proportional allocation both incentivize PMs to inflate orders in an attempt to maximize their own profits and are thus not considered appropriate. While uniform allocations and turn-and-earn policies are robust against gaming, they do not consider relevant parameters such as profitability levels which vary substantially across PMs. In addition, turn-and-earn is especially unsuitable for the problem at hand, as it bases future allocations purely on past sales realizations. The market for CPPs is, however, highly uncertain and single demand realizations are an insufficient basis for future allocation decisions. The case company currently uses a lexicographic allocation policy, which we discuss in detail in Subsection 4.4.2.

4.2.3. Mechanism design

Mechanism design aims to construct environments in which self-interested agents behave in line with the objective of the mechanism designer and in which truthfully disclosing private information is in the agents' best interest (Vohra 2012, Narahari et al. 2009). The literature on mechanism design includes market- or auction-based coordination approaches (McAfee and McMillan 1987, Samuelson 2001, Karabati and Yalçin 2014), as well as various other measures such as penalty schemes, bonuses, transfer pricing, and performance assessments, which aim to encourage truthful information disclosure (Karabuk and Wu 2005). While these mechanisms could theoretically solve the discussed allocation problem, their practical implementation is complex. Coordinating resource allocations based on transfer payments, for example, requires an internal market to determine transfer prices (Vohra 2012, Karabati and Yalçin 2014, Kim et al. 2021). For the problem at hand, a separate transfer price must be derived for each AI, which would require a complex market design such as a combinatorial auction. Determining the transfer prices of each AI through an internal market is deemed too complex by the case company.

Our work contributes to the existing literature by proposing a hybrid allocation approach that leverages concepts from mechanism design, while remaining practical to implement. Our approach incentivizes agents to enhance an initial top-down allocation determined by the HQ, through subsequent bottom-up adjustments in exchange for a transfer payment. Instead of determining appropriate transfer prices through a complex market mechanism, we propose approximating transfer prices by leveraging the limited information available to the HQ, using the multi-tier binary search developed by Zhang (2012).

4.3. Model formulation

In the following, we first provide a formal model representation of the previously introduced allocation problem. We then discuss the generic lower-level decision problem faced by each PM on how to respond to a given allocation mechanism in Subsection 4.3.1, followed by the upper-level problem faced by HQ, namely which allocation mechanism to apply, in Subsection 4.3.2.

We consider a corporation consisting of a central HQ and multiple PMs. The corporation sells $f \in \mathcal{F} = \{1, ..., F\}$ final products. The HQ is responsible for organizational steering, production, and coordinating the distribution of products through the global supply chain network. PMs are responsible for marketing the company's products to customers. Each final product is managed by exactly one PM, and each PM manages exactly one final product. To provide the desired crop protection effects, the corporation produces active ingredients $a \in \mathcal{A} = \{1, ..., A\}$ and blends them into formulations as final products. Each final product requires at least one active ingredient, but could potentially require multiple AIs. We refer to the units of AI a required to produce one unit of final product f as $i_{a,f}$. Available AI quantities are fixed and determined by limited AI production capacities. For the upcoming sales season, c_a units of AI a are available and can be blended into final products to meet market demands. Each final product is market-specific and thus only sold in a single sales market. By deciding how much of each product to blend, the HQ determines the total volumes available to each PM to meet the demand of his respective sales market and allocates the available AI volumes across the corporation.

The demand for CPPs is uncertain. We refer to the uncertain seasonal market demand for final product f as d_f, characterized by a probability distribution f_f with finite supports (d_f^-, d_f^+) and a cumulative distribution function F_f. We assume the lower support is non-negative, i.e., $d_f^- \geq 0$. Because of the limited selling season for CPPs, decisions on how much of each formulation to supply to each sales region must be made before any realization of uncertain market demand. Each unit of final product f sold generates a profit of m_f. Any unit of produced but unsold product must be carried over in local warehouses to the next selling season, which generates handling and storage costs of h_f.

The case company operates a single business planning tool for supply chain and financial planning, which is accessible to both the HQ and the PMs. We therefore assume that per unit profits of product f (m_f), required units of AI a to produce one unit of product f ($i_{a,f}$), available AI quantities (c_a), and warehousing costs (h_f) associated with carrying one unit of unsold product f over into the next selling season are known to both the HQ and the PMs. Conversely, assessing the probability of different demand realizations requires direct market access, which is exclusively held by the responsible PM. We therefore assume that the CDF of uncertain market demand for final product f, $F_f(d_f)$, is only known to the respective PM. Despite lacking direct access to its local markets, the HQ of the case company leverages centrally available market intelligence information on the range of expected crop areas and market shares to estimate worst-case and best-case demands for each market. We assume that these lower and upper bounds correspond to the support of the distribution of uncertain demands. While this assumption simplifies both the analysis and notation in the following, any worst-case and best-case bounds specified by HQ are sufficient and need not coincide with the actual support.

Prior to the start of the annual selling season, HQ must decide how much of each final product to blend from the available AI volumes. By determining the volumes to be blended, HQ allocates the available AI volumes across its final products and determines the inventory of finished goods available to meet the uncertain demand of each market. During this allocation process, HQ applies a predefined and communicated allocation mechanism. The allocation mechanism is a function $\mathcal{M}$ that takes a vector of submitted orders by PMs as input ($\vec{O} = (O_1, \ldots, O_F)$) and returns a vector of confirmed supply quantities ($\vec{Q} = (Q_1, \ldots, Q_F)$), i.e., $\mathcal{M} : \vec{O} \rightarrow \vec{Q}$. Figure 4.1 summarizes the sequence of events in the allocation process.

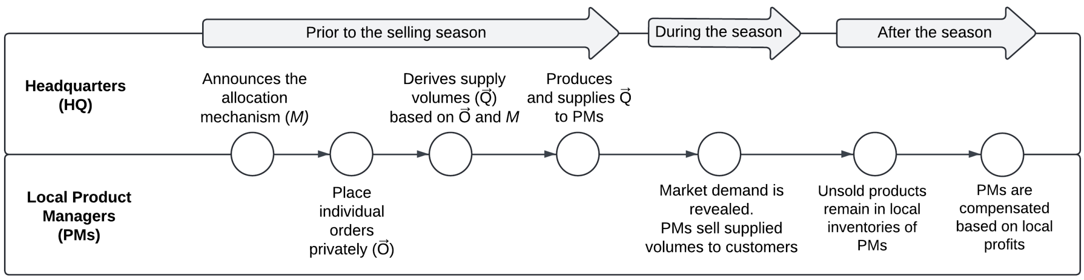

Figure 4.1.: Sequence of events in the allocation process.

In the following section, we first analyze the inner decision problem faced by each PM on how much to order under a given allocation mechanism and then examine the resulting outer decision problem faced by the HQ, namely which allocation mechanism to choose while anticipating the PMs' responses in the inner problem.

4.3.1. Decision problem of the PM: Optimal order quantity in response to a given allocation mechanism

The goal of each PM is to optimize their expected local profits ($E[\pi_f]$), which consist of gross profits from sales to customers minus handling and storage costs generated by holding unsold products in local storage until the next selling season. We denote O_f^s as the self-optimizing quantity that maximizes the expected profits ($E[\pi_f]$) of PM f. This quantity can be determined by solving a traditional single-item NVP, which balances the risk of lost market profits against the risk of additional costs associated with leftover inventory:

$$\underset{O_f^s}{max} \quad E[\pi_f] = m_f \cdot E[\min(d_f, O_f^s)] - h_f \cdot E[(O_f^s - d_f)^+]. \tag{4.1}$$

The self-optimizing quantity O_f^s of PM f can be expressed as:

$$O_f^s = F_f^{-1}(\frac{m_f}{m_f + h_f}). \tag{4.2}$$

4.3.2. Decision problem of the HQ: Definition of allocation mechanism

We now turn to the decision problem faced by HQ regarding which allocation mechanism $\mathcal{M}$ to choose for the upcoming selling season.

As discussed in Subsection 4.1.1, the case company has expressed a clear preference for strategy-proof allocation mechanisms. Consequently, we focus exclusively on such mechanisms in the following. Let $\vec{O}_{i \neq f}$ denote the vector of order quantities of all PMs except PM f. An allocation mechanism $\mathcal{M}$ is strategy-proof if truthfully ordering the self-optimizing quantity is a dominant strategy for all PMs:

$$E[\pi_f(\mathcal{M}_f(O_f^s, \vec{O}_{i \neq f})] \geq E[\pi_f(\mathcal{M}_f(\vec{O})] \quad \forall \vec{O} \in \mathbb{R}^F, \forall f \in \{1, \ldots, F\}. \tag{4.3}$$

We will separately derive and analyze the optimal order policy of PMs in response to each of the allocation mechanisms in the following sections.

In addition to being strategy-proof, the properties outlined in Subsection 4.1.1 specify that any allocation mechanism $\mathcal{M}$ must be implementable and feasible. An allocation mechanism is implementable if it respects the asymmetric information environment presented earlier. An allocation mechanism is feasible if it transforms any order vector

$\vec{O} \in \mathbb{R}^F$ into a vector of supply quantities $\vec{Q}$ that can be produced from the available AI volumes, i.e.,

$$\mathcal{M}(\vec{O} \in \mathbb{R}^F) \rightarrow \vec{Q} : \sum_{f \in \mathcal{F}} i_{a,f} \cdot Q_f \leq c_a \ \forall a \in \mathcal{A} \ \& \ Q_f \geq 0 \ \forall f \in \mathcal{F}. \tag{4.4}$$

In addition to fulfilling the three properties, the HQ aims to design an allocation mechanism $\mathcal{M}$ that maximizes the firm's total expected profits.

4.4. Solution approaches

In the following, we first provide an optimization model to derive the first-best solution of the allocation problem under symmetric information in Subsection 4.4.1. We then present the approach currently used by the case company in Subsection 4.4.2 and subsequently introduce our proposed hybrid allocation approach in Subsection 4.4.3.

4.4.1. First-best solution

We derive the first-best solution of the allocation problem under symmetric information based on a monolithic, integrated optimization model. The problem of determining which final products to produce from the available AI volumes in order to maximize the total expected profits of the company can be formulated as a multi-product NVP with multiple resource constraints:

$$\max_{\vec{Q}} E[\pi] = \sum_{f \in \mathcal{F}} (m_f \cdot E[\min(d_f, Q_f)] - h_f \cdot E[(Q_f - d_f)^+]) \tag{4.5}$$

$$\text{s.t.} \sum_{f \in \mathcal{F}} i_{a,f} \cdot Q_f \leq c_a \qquad \forall a \in \{1, \ldots, A\} \tag{4.6}$$

$$Q_f \geq 0 \qquad \forall f \in \{1, \ldots, F\} \tag{4.7}$$

Objective function (4.5), which is to be maximized, represents the firm's expected total profits. It consists of the sum of expected profits generated from the sale of each final product minus the expected costs generated by holding leftover products in local inventory until the next selling season. Constraints (4.6) ensure that final products can be blended from available AI quantities. Constraints (4.7) enforce that supplied product volumes must be non-negative.

To solve the capacitated multi-product NVP, we apply the approach proposed by Zhang (2012). Using Proposition 1 from Zhang (2012), we introduce $\vec{\lambda} = (\lambda_1, \dots, \lambda_A)$ as a vector of Lagrange multipliers representing the shadow prices of the AI availability constraints (4.6), and reformulate the original optimization problem as follows:

$$\max_{\vec{\lambda}} E[\pi] = \sum_{f \in \mathcal{F}} (m_f \cdot E[\min(d_f, Q_f)] - h_f \cdot E[(Q_f - d_f)^+]) \tag{4.8}$$

$$\text{s.t. } Q_f = F^{-1}\left(\left(\frac{m_f - \sum_a i_{a,f} \cdot \lambda_a}{m_f + h_f}\right)^+\right) \qquad \forall f \in \{1, \dots, F\} \tag{4.9}$$

$$\sum_{f \in \mathcal{F}} i_{a,f} \cdot Q_f \leq c_a \qquad \forall a \in \{1, \dots, A\} \tag{4.10}$$

To solve the reformulated decision problem and derive the optimal allocation quantities, we compute the optimal shadow price vector $\vec{\lambda}$ based on the multi-tier binary search method presented by Zhang (2012). We refer the reader to Zhang (2012) for more details on the solution method employed. In the distributed information environment faced by the case company, the NVP is not implementable. Instead, we use it as a theoretical first-best benchmark in Sections 4.5 and 4.6.

4.4.2. Current practice: Step-wise lexicographic allocation

The case company currently employs a two-step lexicographic allocation mechanism (SLA) to determine how much of the submitted order volume to supply to each PM.

In the first round of SLA, the HQ ensures that the minimum demand of all products is covered. In the second round, the remaining AI volumes are allocated lexicographically up to the full order volume. To determine the sequence in which PMs are served in the allocation process, the HQ ranks all PMs in descending order of their per-unit profitability. The per-unit profitability of product f is derived by dividing the total AI quantity ($\sum_{a \in \mathcal{A}} i_{a,f}$) required to formulate one unit of product f by the margin (m_f) generated per unit sold:

$$\text{Rank}_f = \frac{\sum_{a \in \mathcal{A}} i_{a,f}}{m_f}. \tag{4.11}$$

Rank_f therefore expresses the inverse of the profit per total required AI volume. Note that the ranking is based on information known to the HQ and cannot be influenced by the PMs, i.e., through the submitted orders.

After receiving all orders, the HQ allocates the lower of the minimum market demand and the submitted order quantity to each PM lexicographically. In the second allocation step, the remaining AI volumes are again distributed among PMs by lexicographically confirming the full ordered quantity until either none of the volumes can be increased further due to limited AI availability, or each PM has received his full order.

Since PMs cannot influence the allocation sequence and ultimately receive their submitted order when possible, there is no strategic incentive to deviate from ordering O_f^s. Thus, the optimal order quantity of PM f under SLA is $O_f^{SLA} = O_f^s$ and the mechanism is strategy-proof.

4.4.3. Proposed allocation mechanism: Base-and-refine

One common strategy-proof mechanism to ensure that limited resources are allocated appropriately across agents to maximize the firm's expected profits under private information is to charge agents for claiming access to limited corporate resources through transfer pricing (Vohra 2012). By charging a cost to agents for consuming a shared resource, the central decision maker can ensure that self-interested agents who optimize their own local profits only claim the resource if the expected generated value from its local use exceeds the requested transfer payment.

If the HQ were to charge the shadow prices of the AI availability constraints (4.6) associated with the first-best solution in Subsection 4.4.1 as the transfer price for each respective AI, the resulting self-optimizing quantity of each PM would maximize the firm's expected total profit. Under limited central information availability, the HQ cannot solve the integrated optimization model to determine the optimal shadow prices. One approach to determine appropriate transfer prices under private information is to leverage a market mechanism, such as a combinatorial auction, which is, however, considered too complex for practical implementation by the case company.

Rather than using a market mechanism to determine transfer prices, the HQ can leverage its available information to set transfer prices centrally. This centralized pricing approach faces several problems. If transfer prices are set too low, submitted orders are too high and cannot be served based on the available AI volumes. The HQ must therefore determine how to prioritize the submitted orders to resolve any remaining shortages. If the transfer prices are set too high, orders submitted by the PMs are too low, and available AIs remain unused, resulting in excessive lost sales. We assess the performance impact of deviations between charged transfer prices and true shadow prices in Subsection 4.5.5, and find that expected performance rapidly deteriorates when the charged transfer prices deviate from the true shadow prices.

Instead, we propose a hybrid mechanism, referred to as *base-and-refine*, which combines an initial top-down allocation by the HQ, with the possibility for PMs to exchange initially allocated volumes bottom-up for a centrally defined transfer payment.

The proposed mechanism offers two advantages. First, the transfer payment scheme in the second step offers the PMs both the opportunity and an incentive to jointly correct the initial top-down allocation by leveraging their private information. Any adjustment to the initial top-down allocation will improve the firm's total expected profit, as self-optimizing PMs only request to receive more or fewer volumes if the expected impact on local profits exceeds the associated transfer payment owed or received. Second, by starting with a feasible base allocation and only leveraging the transfer payment to swap selected volumes between PMs in the second allocation step, the performance of our proposed mechanism is substantially more robust to transfer price errors than a pure transfer-pricing-based approach (TP). We assess the latter in more detail in Subsection 4.5.5.

Our proposed approach consists of two steps.

In the first step, the HQ must determine a top-down base allocation ($\vec{B} = (B_1, \ldots, B_F)$) and a transfer price per AI ($\vec{\tau} = (\tau_1, \ldots, \tau_A)$) at which PMs can submit transfer requests in the second step. Both the base allocation and the transfer prices are communicated to the PMs.

In the second step, each PM can submit transfer requests ($\vec{S} = (S_1, \ldots, S_F)$) to the HQ. A positive/negative transfer request indicates that a PM would like to receive more/less than the base volumes. Any confirmed positive/negative transfer request triggers a per-unit transfer payment of $\sum_{a \in \mathcal{A}} i_{a,f} \cdot \tau_a$ to or from the PM, respectively. Each PM submits a single transfer request to the HQ.

Subsequently, the HQ matches positive with negative transfer requests. To decide which PMs to match, we use the lexicographic profitability-based priority list currently used by the case company. The HQ decides which requests to confirm by matching the PM highest up the priority list with an open transfer request to the PM highest on the list with an open opposing transfer request. The process continues until all requests are confirmed or no additional matches can be identified. Subsequently, PMs receive the adjusted allocated quantities to meet market demands. After the end of the business year, the performance of the PMs is assessed based on their realized market profits adjusted by the received/paid transfer charges. We provide further details on each of the three stages below.

Step 1: HQ derives base allocation and sets transfer prices

To determine the base allocation ($\vec{B}$) and derive transfer prices ($\vec{\tau}$), we use the model structure of the first-best solution presented in Subsection 4.4.1 and leverage all information available to the HQ. As full information on the distribution of uncertain market demands is private knowledge held by the responsible local PM, we assume that demand is uniformly distributed between the lower and upper support, which are known to HQ:

$$F_f^U(d_f) \sim U(d_f^-, d_f^+). \tag{4.12}$$

We solve the integrated problem formulation (4.8)-(4.10) based on the problem parameters known to the HQ and the assumption of uniformly distributed demand to obtain the base allocation $\vec{B}$, and we use the shadow prices determined by the solution process of Zhang (2012) as transfer prices $\vec{\tau}$. We refer to the model used to derive the base allocation as IO^U. The HQ communicates both the base allocation and transfer prices to its PMs.

Step 2: PMs submit transfer requests to HQ. HQ determines final allocation quantities

After receiving $\vec{B}$ and $\vec{\tau}$, PMs have the opportunity to submit transfer requests to the HQ. Through transfer requests, PMs have a chance to request volumes above ($S_f > 0$) or below ($S_f < 0$) the base allocation, in exchange for making or receiving a per-unit transfer payment of $\sum_{a \in \mathcal{A}} i_{a,f} \cdot \tau_a$.

After collecting all transfer requests from PMs, the HQ determines which requests to confirm. To do so, the HQ ranks PMs according to the profitability-based lexicographic list used in the company's current allocation approach (see Section 4.4.2). The HQ processes the requests sequentially based on lexicographic priorities, starting with the highest-ranked PM. At each step of the list, the HQ matches the PM with the highest priority and an open request to the next PM on the list with an open opposing request for each of the AIs required by PM f. If a transfer request S_f is confirmed, the base allocation B_f of PM f is updated to $Q_f = B_f + S_f$. Any confirmed transfer generates cross-payments, and the PMs receiving additional volumes pay $(\sum_{a\in\mathcal{A}} i_{a,f} \cdot \tau_a) \cdot S_f$ to the matched PMs. After each match, the request list is updated. The process continues until all requests are met or no further transfers are possible due to the absence of open opposing requests. We provide the full process in Algorithm C.1 in Appendix C.

Under the proposed method, PMs only communicate with the HQ based on the submitted transfer requests. Through these requests, PMs can neither influence the base allocation, the charged transfer prices, nor the process through which transfer requests are matched. If PM f is up for evaluation based on the priority list in the matching phase, the HQ will adjust the base allocation of f and confirm as much of the submitted transfer request (S_f) as allowed by the sum of opposing transfer requests submitted by lower-priority PMs. It is thus in each PM's best interest to truthfully submit the transfer request they would ideally like to have fulfilled in order to maximize their own local expected profits:

$$\underset{S_f}{max} \quad E[\pi_f] = m_f \cdot E[\min(d_f, (B_f + S_f)] - h_f \cdot E[(B_f + S_f - d_f)^+] - \sum_{a\in\mathcal{A}} \tau_a \cdot i_{a,f} \cdot S_f \tag{4.13}$$

$$\text{s.t.} \quad B_f + S_f \geq 0 \tag{4.14}$$

which leads to the following optimal transfer request:

$$S_f = F^{-1}((\frac{m_f - \sum_{a\in\mathcal{A}} \tau_a \cdot i_{a,f}}{m_f + h_f})^+) - B_f. \tag{4.15}$$

If the transfer prices set by the HQ equal the true shadow prices obtained from the integrated model formulation, PMs will correct any base allocation to a final allocation that leads to the same expected total profits as the first-best solution obtained in the absence of private information in Subsection 4.4.1. BR fulfills the three properties desired by the case company. As discussed, BR is strategy-proof because PMs have an incentive to truthfully reveal their self-optimizing preferences when submitting transfer requests to the HQ. Furthermore, BR is both feasible and implementable, as the initial base allocation and any bottom-up adjusted volume allocations solve the allocation problem, and the mechanism can be implemented within the decision-making context of the case company.

4.5. Numerical study based on synthetic data

In this section, we evaluate the performance of our proposed allocation method (BR) against the company's current practice (SLA), additional benchmarks from theory and practice, as well as the first-best solution achievable under full central information availability based on the integrated model formulation (IO). Using a synthetic data set with systematically varied input parameters, we aim to analyze the performance and solution structure of our proposed allocation mechanism. The study setup is detailed in Subsection 4.5.1, the benchmark methods are formalized in Subsection 4.5.2, and the performance analysis is presented in Subsection 4.5.3 onward.

4.5.1. Study setup

In our numerical study, we consider two AIs from which a portfolio of 12 final products can be produced. The final products are blended from either one or both AIs. Each product is managed by one dedicated PM. To evaluate the allocation methods, we construct 12 different portfolios, each containing 12 final products. To construct the portfolios, we differentiate between parameters held constant across all portfolios (see Table (4.1)) and parameters that are varied (see Table (4.2)).

We assume that each final product has an expected demand ($E[d]$) of 1000 units. As all studied allocation methods adjust allocation results proportionally to level shifts in expected demand, studying the impact of different market sizes on allocation outcomes is not relevant. Carrying leftover inventory to the next selling season generates additional local handling and storage costs (h_f) of 1.5 per unit. In addition, we assume that each unit of final product requires a total of 0.2 units of AI. The portfolio consists of final products that can be produced from one of three different AI consumption vectors ($\vec{i}_f$), requiring either 0.2 units of AI 1, 0.2 units of AI 2, or 0.1 units of both AI 1 and AI 2. The number of mixed AIs and the AI density realistically reflect the composition of CPPs currently sold in the market. We summarize the fixed parameters used to construct the product portfolio in Table 4.1.

Our preliminary studies reveal that allocation outcomes are strongly sensitive to the skewness of assumed market demand distributions (γ), as well as to the degree of diversity in demand uncertainty (cv) and profitability levels (m) across final products. Thus, we design 12 distinct portfolios, combining left-, right-, or non-skewed demand distributions with both homogeneous and heterogeneous demand uncertainties and profitability levels, as detailed in Table 4.2. The ratios of the different profitability levels to the holding cost parameter are realistically aligned with those observed in the industry.

Parameter	**Value**
$E[d_f]$	1000
h_f	1.5
$\vec{i}_f$	$\{0.2, 0.0\}, \{0.0, 0.2\}, \{0.1, 0.1\}$

Table 4.1.: Problem parameters not varied across portfolios.

Parameter	**Low diversity**	**High diversity**
cv	{0.17, 0.23}	{0.1, 0.3}
m	{7, 9}	{4, 12}
γ	$\{-0.995, 0.000, 0.995\}$	

Table 4.2.: Problem parameters varied across portfolios.

We model uncertain market demand using a skew normal distribution, defined by the given coefficient of variation (cv_f) and skewness (γ_f), truncated at the 1^{st} and 99^{th} percentiles.

For each portfolio, we parameterize the 12 final products based on all possible combinations of the three different AI consumption vectors ($\vec{i}_f$) in Table 4.1 and the respective values of high or low diversity in cv_f and m_f from Table 4.2 such that $3(\vec{i}_f) \cdot 2(cv_f) \cdot 2(m_f) = 12$ final products.

We determine allocations for each of the 12 portfolios across three AI availability scenarios: both AIs with high availability (H/H), one AI with high and the other with low availability (H/L), and both AIs with low availability (L/L). 'High' and 'low' AI availability is defined relative to the total volumes required to satisfy the unconstrained, self-interested orders of all PMs (O^s_f), with high availability set at 90% and low availability at 70% of the required AI volumes.

4.5.2. Benchmark methods

We evaluate the performance of the proposed allocation method (BR) by comparing it to several benchmark strategies outlined in Table 4.3. In addition to the proposed allocation mechanism (BR), the current company practice (SLA), and the first-best solution under full central information availability (IO), we also test the performance of the base allocation $\vec{B}$ used in BR, which is derived by using the integrated problem formulation under the assumption of uniformly distributed market demands (IO^U). Furthermore, we include several additional benchmark methods that are either well-established in the literature or currently under consideration by the case company. These methods are presented below.

Proposed Method	**Description**
BR	Base-and-refine.
Theoretical Benchmark	**Description**
IO	First-best solution obtainable under full central information availability.
LA	Traditional lexicographic allocation method.
IO^U	Base allocation used by BR and derived through integrated model under the assumption of uniformly distributed demand.
TP	Allocation based on traditional transfer pricing.
In-company Benchmark	**Description**
SLA	Step-wise lexicographic allocation. Current company practice.
UPA	Uniform-proportional allocation. Under discussion at case company to potentially improve SLA.

Table 4.3.: Summary of studied allocation methods.

Note that all allocation methods face the challenge of imposing the right level of prioritization among the articles in the company's portfolio. If volumes are too aggressively allocated toward the most profitable products, the case company might experience high leftovers in high-profit products while dealing with shortages in products with lower profitability. If volumes are, however, allocated too equally, the company might increase sales of low-profit products but simultaneously lose substantial profits from articles with the highest profitability. The methods studied span a broad range of prioritization strategies. While LA, for example, prioritizes volume allocations purely based on profitability, UPA allocates volumes uniformly across all PMs and fully disregards relative differences in profitability.

Lexicographic Allocation (LA)

While the case company currently employs a two-step lexicographic allocation approach, we include the traditional single-step lexicographic approach (LA) as an additional benchmark in the following analysis (Cachon and Lariviere 1999b). By including LA, we enable a direct comparison between the company's current practice and a closely related, well-established allocation method.

Allocation mechanism applied by the HQ:

Under LA, the HQ uses the same profit-based ranking of PMs as in SLA (see equation (4.11)) to determine the sequence in which PM orders are fulfilled. LA is therefore both implementable and feasible.

Optimal order strategy of PMs under allocation mechanism:

Similar to SLA, LA provides no incentives for PMs to deviate from ordering the unconstrained, self-optimizing quantity (Cachon and Lariviere 1999b). Thus, the optimal order policy for PMs under LA is

$$O_f^{LA} = O_f^s \tag{4.16}$$

and LA is strategy-proof.

Allocation based on transfer prices (TP)

In the second step of our proposed allocation method, we suggest leveraging the concept of transfer prices to improve a top-down base allocation provided by the HQ. To assess the performance of our approach, we compare it to the performance of a traditional pure transfer price-based mechanism (TP) commonly found in the literature.

Allocation mechanism applied by the HQ:

Under TP, the HQ first announces a vector of transfer prices ($\vec{\tau}$) to its PMs. As in BR, we set transfer prices based on the shadow prices obtained when determining the base allocation $\vec{B}$ based on IO^U. Subsequently, PMs submit orders to the HQ. If all orders can be blended from the available AI volumes, PMs receive their full orders. If required AI quantities exceed available AI volumes, the HQ performs a lexicographic allocation based on LA on the submitted orders. The HQ blends and supplies the final confirmed volumes to the PMs in exchange for the associated transfer payment. As TP relies solely on centrally available information and fully resolves any allocation conflict, the method is implementable and feasible.

Optimal order strategy of PMs under allocation mechanism:

As with the other lexicographic-based mechanisms, it remains optimal for PMs to order the self-optimizing quantity while accounting for the transfer payment due, in order to maximize expected local profits. The self-optimizing quantity of a PM under $\vec{\tau}$ can be characterized by the newsvendor trade-off presented in equation (4.9). TP is therefore strategy-proof.

Uniform-proportional allocation (UPA)

By using SLA, the case company currently generates allocation plans in which the most profitable PMs receive their full, self-optimizing, unconstrained order quantities, while product supply for other PMs is reduced to at most the lower bound of uncertain demand. This has resulted in multiple years in which some PMs held substantial leftover stocks, while others simultaneously failed to capture significant market demand. Prior to engaging in the research collaboration, the case company therefore considered implementing an allocation mechanism that distributes volumes more uniformly across its PMs. We present the approach under discussion, referred to as uniform-proportional allocation (UPA), together with the optimal order policy of PMs, and assess its performance as an additional benchmark.

Allocation mechanism applied by the HQ:

Under UPA, all PMs initially place an order for final products with the HQ. Subsequently, the HQ offers each PM requiring a given AI their certain market demand (d_f^-) plus the same share (s_a) of their uncertain demand component ($d_f^+ - d_f^-$). Note that the HQ, in contrast to the uniform allocation discussed in the literature review, does not offer volumes proportional to submitted orders, but based on the range of uncertain demand derived from centrally available information. In cases of severe allocation constraints, where not even the lower bound of uncertain demand for final products can be supplied, each PM is offered the same share s_a of the lower bound of their market demand. If a final product is blended from more than one AI, the offered volume corresponds to the lowest quantity determined across the required AIs. Finally, PMs receive the minimum of the offered volumes and their submitted orders.

We therefore define $Q_{f,a}$ as

$$Q_{f,a}(s_a) = \begin{cases} min\{d_f^- + s_a \cdot (d_f^+ - d_f^-),\, O_f\} & \text{if } \sum_{f\in\mathcal{F}} i_{a,f} \cdot \min(O_f, d_f^-) \leq c_a \\ min\{s_a \cdot d_f^-,\, O_f\} & \text{if } \sum_{f\in\mathcal{F}} i_{a,f} \cdot \min(O_f, d_f^-) > c_a \end{cases} \tag{4.17}$$

where the maximal feasible share s_a suppliable for final products requiring AI a is defined as

$$\underset{s_a}{max}\{s_a : \sum_{f\in\mathcal{F}} i_{a,f} \cdot Q_{f,a}(s_a) \leq c_a\}. \tag{4.18}$$

The final allocated quantity Q_f^{UPA} is the lowest AI-specific quantity $Q_{f,a}$ derived for final product f:

$$Q_f^{UPA} = \underset{a\in A,\, i_{a,f}>0}{min}\{Q_{f,a}\}. \tag{4.19}$$

Any leftover AI volumes after determining supply quantities based on equation (4.19) are distributed to final products requiring only this single AI by proportionally increasing the supplied share of those products up to the submitted orders. In line with the desired properties defined by the case company, SLA can be implemented in the context of distributed information and fully eliminates AI shortages. SLA is thus implementable and feasible.

Optimal order strategy of PMs under allocation mechanism:

Under UPA, PMs receive the lower of the quantity derived based on the uniform share applied by the HQ and the submitted order. The share-based quantity is determined by the HQ based on centrally available information on the lower and upper bounds of uncertain market demands. PMs cannot increase their allocation share through strategic order placement. However, by truthfully reporting their self-optimizing quantities, PMs can ensure that the HQ does not oversupply them and avoid high expected leftover costs in case of high allocation shares. UPA is therefore strategy-proof and PMs will order

$$O_f^{UPA} = O_f^s. \tag{4.20}$$

4.5.3. Performance analysis of allocation methods

We assess the performance of the different allocation mechanisms based on the expected total allocation impact (TAI). $TAI^{\mathcal{M}}$ quantifies the expected total profit impact of limited AI availabilities by subtracting the expected total profits of the corporation resulting from allocating volumes based on mechanism $\mathcal{M}$ in a given AI availability scenario, $\pi(\vec{Q}^{\mathcal{M}})$, from the expected total profits if all PMs receive their unconstrained, self-optimizing quantities in the absence of limited AI availabilities, $\pi(\vec{O}^s)$:

$$TAI^{\mathcal{M}} = \pi(\vec{O}^s) - \pi(\vec{Q}^{\mathcal{M}}). \tag{4.21}$$

During the research collaboration with the case company, using TAI as the performance metric has allowed for two discussions. First, we can use the absolute TAI of the first-best solution to assess the expected profit loss under different availability scenarios. This offers the case company an important perspective on the isolated impact of AI availability constraints on expected profits. The analysis provides the case company a first intuition on how much to invest in relaxing AI availability constraints by, for example, expanding AI production capacities. Second, we can use the performance gap between the first-best solution and the studied allocation methods to quantify the additional allocation impact on expected total profits resulting from limited central information and incentive misalignment when applying different allocation mechanisms. As the objective of this paper is to propose an allocation mechanism that performs well in the presented problem environment, we focus on the latter in the following analysis.

We present the expected TAI when applying our proposed allocation mechanism and the introduced benchmarks, including the case company's current practice, to the 12 different portfolios under three availability scenarios as defined in Section 4.5.1. We summarize the overall results in Table 4.4. In addition to reporting the average performance across all portfolios and availability scenarios, we filter the results by AI availability scenario, diversity in the coefficient of variation of uncertain market demands and profitability levels, and the assumed distribution shape of uncertain demands. The percentage values in parentheses indicate the gap to the TAI of the first-best solution obtainable under full central information (IO).

		First-best	Proposal	Theoretical benchmark			In-company benchmark	
		IO	***BR***	***LA***	***IO^U***	***TP***	***SLA***	***UPA***
	Avg:	**49.7**	**52.8 (6.3%)**	**124.2 (149.9%)**	**56.2 (13.1%)**	**81.5 (64.0%)**	**98.9 (98.9%)**	**63.7 (28.2%)**
Avail	L/L	87.0	89.8 (3.2%)	179.6 (106.6%)	96.0 (10.4%)	117.1 (34.6%)	146.0 (67.9%)	105.7 (21.5%)
	H/L	52.8	56.1 (6.2%)	142.1 (168.8%)	59.6 (12.9%)	82.1 (55.5%)	106.1 (100.7%)	71.1 (34.5%)
	H/H	9.3	12.5 (35.2%)	50.9 (448.7%)	13.0 (40.1%)	45.2 (387.6%)	44.5 (379.9%)	14.4 (55.1%)
Div *cv*	L	49.1	52.3 (6.5%)	121.4 (147.2%)	55.6 (13.3%)	82.5 (68.0%)	100.7 (105.0%)	62.9 (28.0%)
	H	50.3	53.3 (6.1%)	126.9 (152.5%)	56.8 (13.0%)	80.5 (60.1%)	97.0 (93.0%)	64.6 (28.5%)
Div *m*	L	55.3	56.4 (1.9%)	156.6 (183.1%)	56.5 (2.1%)	98.0 (77.2%)	119.8 (116.6%)	58.7 (6.1%)
	H	44.1	49.3 (11.8%)	91.7 (108.2%)	56.0 (27.1%)	64.9 (47.4%)	77.9 (76.8%)	68.7 (56.0%)
Shape	L Skew	61.0	62.0 (1.8%)	126.2 (106.9%)	62.4 (2.4%)	91.2 (49.6%)	110.4 (81.1%)	79.1 (29.7%)
	Symmetric	47.9	49.5 (3.4%)	123.7 (158.5%)	52.0 (8.7%)	55.3 (15.5%)	100.2 (109.5%)	61.3 (28.1%)
	R Skew	40.3	46.9 (16.5%)	122.7 (204.8%)	54.2 (34.7%)	98.0 (143.4%)	86.0 (113.5%)	50.8 (26.2%)

Table 4.4.: Expected total allocation impact across portfolios by availability scenario and portfolio characteristics.

Overall, BR-based allocations lead to a TAI that is 6.3% higher than the first-best solution and outperform all other benchmark methods. Among the benchmarks, the base allocation IO^U (13.1%) and UPA (28.2%) perform best, followed by TP (64.0%). Both methods based on pure lexicographic prioritization perform the worst. Specifically, the current company practice SLA results in an average TAI that is 98.9% higher than that of IO, while LA shows a gap of 149.9%. The results confirm the benefit of leveraging all centrally available information (IO^U) and demonstrate that offering PMs the opportunity to adjust a top-down allocation in exchange for a transfer payment (BR) outperforms the unadjusted base allocation. Note that although BR combines IO^U and the concept of transfer prices behind TP, the performance of BR aligns much more closely with IO^U, as the base allocation determined by IO^U is only selectively adjusted based on bottom-up volume transfers in the second step to enhance overall performance. In addition, the results highlight that prioritizing the allocation of volumes purely based on profitability while ignoring other relevant factors is insufficient. This is evident in the poor performance of LA, which applies a purely profit-based prioritization, and SLA, which follows a solely profitability-driven ranking in the second allocation step. Conversely, incorporating even limited information about market uncertainty (as in UPA) proves beneficial and enhances performance.

The aforementioned effects depend on the structure of the studied portfolio. We find that methods based on lexicographic priorities (SLA, LA, and TP) perform substantially worse under scenarios of high availability. Under these methods, AI shortages are resolved by cutting the demand of a few selected PMs while fully supplying the remaining PMs. Particularly in high AI availability scenarios, this approach results in outcomes where selected PMs experience significant lost sales, whereas fully served PMs incur high expected leftovers. This finding confirms the case company's concern about the performance imbalance caused by fully satisfying the most profitable PMs' orders while other PMs face severe shortages. Employing an allocation method that achieves an appropriate level of prioritization is therefore essential.

We observe that the degree of diversity in the coefficient of variation of uncertain market demands across PMs has no significant impact on the overall performance of the studied allocation methods. In contrast, diversity in profitability levels has a stronger influence on performance across the different methods. BR performs best when profitability diversity is low. When profitability diversity is high, the base allocation (IO^U) performs worse due to its assumption of uniformly distributed demand, which leads to increased over-allocation to the most profitable PMs. BR can only partially mitigate this initial over-allocation through volume swaps (reducing the gap from 27.1% to 11.8%). The performance of all methods that allocate (partially) based on lexicographic priorities (SLA, LA, TP) improves under high profitability diversity. Conversely, the performance of UPA deteriorates as profitability diversity increases.

Regarding the distribution shape of market demands, BR consistently outperforms other mechanisms on average across all distribution shapes. However, the gap to the first-best solution increases from 1.8% for left-skewed to 16.5% for right-skewed demand distributions. This widening gap is primarily driven by the performance limitations of the base allocation (IO^U), which BR can only partially correct through volume transfers. If true demands are right-skewed, assuming uniformly distributed demands overestimates the probability of high market-demand scenarios, which in turn leads to increased over-allocation to the most profitable PMs, driven by high optimal service levels determined by the critical fractile in equation (4.9).

		First-best	Proposal	Theoretical benchmark			In-company benchmark	
		IO	***BR***	***LA***	***IO^U***	***TP***	***SLA***	***UPA***
	Avg:	**37.4**	**48.5 (29.4%)**	**91.7 (145.0%)**	**63.9 (70.6%)**	**76.9 (105.4%)**	**82.7 (120.8%)**	**56.4 (50.6%)**
Avail	L/L	65.9	75.2 (14.1%)	131.7 (99.8%)	105.8 (60.4%)	96.4 (46.2%)	122.0 (85.0%)	91.4 (38.6%)
	H/L	40.3	50.5 (25.4%)	106.6 (164.7%)	66.2 (64.5%)	74.5 (84.9%)	102.4 (154.4%)	62.5 (55.3%)
	H/H	6.1	19.7 (221.5%)	36.9 (502.0%)	19.7 (221.5%)	59.8 (877.5%)	23.6 (286.2%)	15.3 (149.2%)

Table 4.5.: Expected total allocation impact for a portfolio with high variability in profitability and uncertainty under right-skewed demand distributions by availability scenario.

In addition to the aggregated results across all studied portfolios, we present the outcomes for a single portfolio characterized by right-skewed demand distributions and high diversity in both profitability and uncertainty levels among final products in Table 4.5. This portfolio most closely reflects the real-world data of the case company. Again, BR outperforms the other benchmarks (29.4%), followed by UPA (50.6%) and IO^U (70.6%). TP, SLA, and LA perform worse, with 105.4%, 120.8%, and 145.0%, respectively. These results underscore both the superior performance of BR compared to the studied benchmarks and the benefits of allowing PMs to adjust a top-down base allocation through volume transfers, as evidenced by the significant improvement from 70.6% (IO^U) to 29.4% (BR).

The gap between IO and BR widens in high-availability scenarios. Specifically, when the availability of both AIs is high, BR (221.5%) performs even worse than UPA (149.2%). This large percentage deviation stems from the low absolute value of TAI for IO. However, the performance gap for BR is primarily driven by two factors: the limitations of the base allocation and the estimation error between the charged transfer prices and the true shadow prices. When the true demand distribution is right-skewed, the base allocation, assuming uniform demand, tends to over-allocate AIs to the most profitable PMs at the expense of those with lower profitability levels. Simultaneously, the resulting transfer prices are overestimated, discouraging PMs from taking on additional volumes to correct the initial allocation. The consequences of overestimated transfer prices are also evident in the poor performance of TP. When charged transfer prices are set too high, a share of the available AIs remains with HQ, leading to a TAI substantially exceeding the optimal allocation under IO (877.5%). These results highlight the risks of determining allocations solely based on a transfer pricing scheme. We discuss the impact of transfer pricing errors on overall allocation performance in the following section.

4.5.4. Solution structure and driver analysis

Besides the performance analysis, we now investigate the structure of the obtained solutions. Specifically, we investigate how differences in demand uncertainty (cv) and profitability (m) across products translate into differentiated volume allocations across the different mechanisms.

To this end, we consider the portfolios studied in Subsection 4.5.3 under both low and high availability scenarios for both AIs. For each availability scenario, we calculate the resulting average supply quantity of products with low and high coefficients of variation and margins for each allocation method. Under each availability scenario, an identical amount of AI is allocated across products by each allocation mechanism. We can therefore simplify the analysis and focus solely on the interval length and the direction of differentiation between average supply quantities of products with low and high parameter realizations. Accordingly, Table 4.6 reports the percentage difference between the average supply quantity of high versus low parameter realizations by availability scenario and allocation method. Under a low availability scenario, the first-best solution, for example, allocates 8.5% less to products with high demand uncertainty than to those with low demand uncertainty.

		First-best	**Proposal**	**Theoretical benchmark**			**In-company benchmark**	
	Avail	IO	BR	LA	IO^U	TP	SLA	UPA
cv	L/L	-8.5%	-8.1%	24.1%	-4.1%	-8.5%	-11.7%	-6.5%
	H/H	11.3%	12.4%	3.8%	12.9%	0.0%	7.6%	12.9%
m	L/L	36.9%	42.7%	189.7%	60.6%	57.3%	147.8%	0.0%
	H/H	14.0%	24.7%	37.1%	25.9%	24.3%	37.1%	0.0%

Table 4.6.: Percentage difference in average supply quantities between products with high and low coefficients of variation and profitability by AI availability scenario.

We can determine the optimal degree and direction of differentiation by studying the allocations obtained through the first-best solution. We find that if the demand uncertainty of a product is high, it is optimal to allocate lower volumes under low availability scenarios (-8.5%) and higher volumes under high availability scenarios (11.3%). If available AI volumes are scarce, it is optimal to prioritize the supply of products with more certain demands. If AI volumes are more abundant, allocating additional volumes to high-uncertainty markets becomes optimal to capture a larger share of uncertain demand. Intuitively, it is optimal to allocate higher volumes to more profitable products under both availability scenarios. The relative degree of differentiation is stronger under the low availability scenario (36.9%) compared to the high availability scenario (14.0%).

We can use the results of the first-best solution to assess how closely the allocation mechanism proposed by this paper and the studied benchmarks follow the optimal differentiation pattern. We find that for the degree of demand uncertainty, BR, IO^U, SLA, and UPA follow the pattern of IO and allocate less to products facing high demand uncertainty under a low AI availability scenario while allocating more to high-uncertainty products when AI availability is high. Interestingly, LA is the only method that allocates substantially higher volumes to products facing high demand uncertainty under a low availability of AIs. Under LA, orders of PMs are strictly prioritized based on per-unit margins. If the availability of AIs is low, the majority of PMs receive no volumes, while a small number of the most profitable PMs receive their full, unconstrained orders. The optimal self-optimizing order quantity of PMs can be determined based on the NVP (see equation (4.2)). For the most profitable products, the critical fractile is above 0.5 for the studied data, which leads PMs with high demand uncertainty to order more than those with lower demand uncertainty.

When examining differentiation patterns based on profitability, all methods except UPA allocate more to PMs with higher per-unit margins under both availability scenarios. As for IO, the degree of differentiation is lower under a high availability scenario. LA and SLA prioritize orders of PMs purely based on profitability, which leads to over-differentiation when compared to the degree of differentiation resulting from the first-best solution. UPA only allocates volumes based on the degree of demand uncertainty and does not consider differences in profitability across PMs. Overall, we find that our proposed method follows both the optimal direction and the degree of differentiation of the first-best solution more closely and consistently than any other benchmark method.

4.5.5. Impact of shadow price estimation errors on expected performance

The proposed allocation mechanism (BR) and the benchmark method based on a pure transfer pricing scheme (TR) rely on transfer prices centrally set by the HQ. If the HQ had full information, it would ideally charge the true shadow price of each AI-availability constraint, as determined in the first-best solution, to each PM in order to fully align the maximization of the firm's expected total profits with the incentives of PMs to maximize their local expected profits. Due to private information held by each PM, the HQ must set transfer prices by approximating true shadow prices based on limited central information (see step 1 in Section 4.4.3).

In the following, we first assess the approximation error between the true shadow prices and the approximated transfer prices, as reported in Table 4.7, and subsequently evaluate the performance impact of these estimation errors on the overall performance of both methods, as shown in Table 4.8 and Table 4.9.

		λ_1^{IO}	λ_1^{U}	λ_2^{IO}	λ_2^{U}
	Avg:	**12.5**	**13.5 (7.5%)**	**12.5**	**13.5 (7.5%)**
Avail	L/L	19.0	17.6 (-7.3%)	19.0	17.6 (-7.3%)
	L/H	21.8	19.4 (-10.7%)	3.2	7.5 (136.5%)
	H/L	3.2	7.5 (136.5%)	21.8	19.4 (-10.7%)
	H/H	6.1	9.2 (50.7%)	6.1	9.2 (50.8%)
Div cv	L	12.4	12.8 (3.5%)	12.4	12.8 (3.5%)
	H	12.7	14.1 (11.3%)	12.7	14.1 (11.3%)
Div m	L	14.0	15.3 (9.0%)	14.0	15.3 (9.0%)
	H	11.0	11.6 (5.5%)	11.0	11.6 (5.5%)
Shape	L Skew	14.3	6.5 (-54.7%)	14.3	6.5 (-54.7%)
	Symmetric	12.3	13.4 (9.6%)	12.3	13.4 (9.6%)
	R Skew	11.0	20.4 (86.4%)	11.0	20.4 (86.3%)

Table 4.7.: Comparison of shadow prices obtained under IO and IO^U across portfolio characteristics and availability scenarios.

In Table 4.7, we assess the average true shadow price (λ_i^{IO}) together with the estimated transfer price based on the shadow prices obtained when solving IO^U (λ_i^U), which are derived from centrally available information under the assumption of uniformly distributed demands. We indicate the percentage gap between the true and approximated shadow prices in parentheses. Across all studied availability scenarios and portfolios, the estimated shadow price exceeds the true shadow price on average by 7.5%. However, this overall average is not an appropriate measure, as the approximation error strongly depends on the problem parameters, which we examine in the following.

Under low availability scenarios (L/L), IO^U underestimates true shadow prices by 7.3%. This occurs because the uniform distribution overestimates the likelihood of very low demand, thereby underestimating the marginal profit impact of additional AI volumes under low availability scenarios. Conversely, in high availability scenarios (H/H), IO^U overestimates true shadow prices by 50.7%, as it overestimates the probability of high demand realizations.

Compared to differences across availability scenarios, shadow price errors exhibit only low sensitivity to the degree of diversity in the coefficient of variation of uncertain demands (3.5% vs. 11.3%) and profitability levels (9.0% vs. 5.5%). However, the skewness of the demand distribution has a stronger impact on the approximation errors. When the true demand distribution is left-skewed, shadow prices estimated based on the assumption of uniform demand are 54.7% lower than the true shadow prices. Conversely, when the true demand distribution is right-skewed, estimated shadow prices exceed the true values by 86.4%. Similar to the discussion on availability levels, the error arises from the uniform distribution's systematic under- and overestimation of demand at risk due to limited availabilities, depending on whether the true demand distribution is left- or right-skewed.

The results in Table 4.7 demonstrate that the accuracy of the approximation of true shadow prices is influenced by the specific parameter characteristics of the problem. To evaluate the robustness of the proposed approach and the traditional transfer pricing method with respect to shadow price estimation errors, we analyze their impact on the expected TAI. To this end, we quantify the expected TAI under varying shadow price approximation errors for both AIs. We present the percentage gap of the resulting expected TAI to the expected TAI of the first-best solution for the traditional transfer scheme TP in Table 4.8 and for BR in Table 4.9.

		Error τ_2				
		-40%	-20%	0%	+20%	+40%
Error τ_1	-40%	49.3%	37.5%	32.8%	61.8%	94.5%
	-20%	37.5%	19.8%	14.0%	44.6%	78.8%
	0%	32.8%	14.0%	**0.0%**	36.8%	78.2%
	+20%	61.8%	44.6%	36.8%	77.0%	121.5%
	+40%	94.5%	78.8%	78.2%	121.5%	161.0%

Table 4.8.: Impact of shadow price estimation errors on expected TAI compared to the first-best solution under TP.

If estimated shadow prices match true shadow prices, the expected performance of both BR and TP aligns with the performance of the first-best solution IO. As errors increase, the expected allocation impact worsens for both methods. For every error combination, the TAI increase of BR (Table 4.9) is significantly below the TAI increase under TP (Table 4.8). BR is therefore substantially more robust to deviations of transfer

		Error τ_2				
		-40%	-20%	0%	+20%	+40%
Error τ_1	-40%	4.7%	3.9%	3.3%	7.5%	10.6%
	-20%	3.9%	2.9%	2.0%	6.4%	9.9%
	0%	3.3%	2.0%	**0.0%**	5.5%	8.4%
	+20%	7.5%	6.4%	5.5%	8.6%	12.4%
	+40%	10.6%	9.9%	8.4%	12.4%	13.2%

Table 4.9.: Impact of shadow price estimation errors on expected TAI compared to the first-best solution under BR.

prices from true shadow prices than TP, consistently performing better across all error scenarios. For both methods, setting transfer prices too high has a stronger negative impact on expected performance than setting them too low.

Under TP, understated transfer prices lead to orders whose AI requirements exceed available AI volumes. Any remaining AI shortage must be resolved in an additional step, for example, by applying the lexicographic priority list of LA. As underestimations increase, TP outcomes converge to those of LA. Conversely, when shadow prices are overestimated, submitted orders can be served based on available AI volumes, but an increasing portion of AI remains unused at the HQ, resulting in substantial profit losses.

The expected TAI of BR demonstrates significantly greater robustness to errors in charged transfer prices. Under BR, transfer prices are used to incentivize PMs to adjust the initial base allocation based on their private information in exchange for a transfer payment. If charged transfer prices are too low/too high, it becomes less attractive for PMs to offer/purchase additional volumes from other PMs. As a result, the volume transferred decreases as estimation errors increase, and the BR solution gradually converges to the base allocation.

4.6. Application to real-world data for the case company

In the following section, we complement the findings based on synthetic data by assessing the performance of the proposed allocation mechanism within the real-world context of the case company.

4.6.1. Data collection and scope of analysis

To assess the performance of the proposed mechanism alongside the discussed benchmarks, we use real-world data from the product portfolio of two key AIs. Both AIs are of special interest to the case company because the associated portfolios generate a substantial share of the company's revenues and profits. For both AIs, the requests of PMs could not be fulfilled based on available AI volumes in several previous selling seasons. The company produces a total of 119 different final products from both AIs. Of these, 86 final products require only one of the AIs, while the remaining 33 require both AIs.

For each final product, we collect the relevant per-unit profits, coefficients of required AI quantities, and local handling and storage costs for leftover products. The probabilities of different demand scenarios for future selling seasons are not known to the HQ of the case company. For the sake of the computational study, we assume that market uncertainties can be quantified based on historical forecast errors. As the lifespan of CPPs is limited, we pool historic forecast errors for similar products sold in similar markets together. This approach allows us to approximate a distribution that represents the demand uncertainty of an average historic selling season, which we use to evaluate the general performance of the different allocation methods. To solve the actual allocation problem of a future selling season, this historic approximation is of limited value because the probabilities of different demand scenarios for a single future season vary with highly volatile and rapidly evolving local farming environments, which remain private knowledge held by the local PMs.

To define the different availability scenarios, we first determine the AI volumes required to fulfill the unconstrained orders of all PMs. We then define 70% of the unconstrained AI requirement as the "low" and 90% as the "high" availability scenario.

4.6.2. Performance analysis

In Table 4.10, we compare the performance of the studied allocation mechanisms under different availability scenarios using real-world data from the case company. BR outperforms all other benchmark methods with an average gap to the first-best solution of 29.0%. Among the benchmark methods, UPA performs best (55.4%), followed by IO^U (87.7%), LA (90.8%), SLA (91.9%), and TP (103.1%). The overall average findings, as well as those for the different availability scenarios, are consistent with the results obtained for the portfolio characterized by high diversity in profitability and demand uncertainty with right-skewed demand distributions, as presented in Table 4.5.

		First-best	Proposal	Theoretical benchmark			In-company benchmark	
		IO	***BR***	***LA***	***IO^U***	***TP***	***SLA***	***UPA***
	Avg:	**92.6**	**119.5 (29.0%)**	**176.6 (90.8%)**	**173.8 (87.7%)**	**188.1 (103.1%)**	**177.7 (91.9%)**	**143.9 (55.4%)**
Avail	L/L	146.8	184.9 (26.0%)	257.2 (75.3%)	236.3 (61.0%)	256.2 (74.6%)	253.2 (72.5%)	230.7 (57.2%)
	L/H	93.0	121.1 (30.2%)	208.1 (123.7%)	164.2 (76.5%)	180.0 (93.5%)	209.9 (125.6%)	126.7 (36.2%)
	H/L	116.0	141.7 (22.1%)	189.0 (62.9%)	199.0 (71.5%)	208.7 (79.8%)	199.2 (71.6%)	188.7 (62.6%)
	H/H	14.6	30.2 (106.9%)	52.4 (259.1%)	95.8 (557.2%)	107.5 (637.7%)	48.6 (233.5%)	29.7 (103.7%)

Table 4.10.: Expected total allocation impact by availability scenario for company data.

The results underline the performance improvement potential when switching from the current company practice, SLA, to the proposed allocation method, BR. If the complexities associated with introducing transfer payments are a concern for successful short-term implementation, UPA offers significant improvement potential, even when employed as an intermediate step.

In addition, the results confirm the strength of BR in allowing PMs to correct and improve an initial, top-down base allocation (IO^U) through subsequent volume transfers. To capture this potential, the case company must, however, establish processes that enable PMs to offer or request volumes during the transfer phase, and to track all transactions appropriately. To this end, it is important to assess the extent to which trades are expected to occur.

		Transferred volumes
	Avg:	**6.7%**
Avail	L/L	5.6%
	L/H	7.2%
	H/L	6.1%
	H/H	7.6%

Table 4.11.: Proportion of total product volume reallocated through bottom-up adjustments in BR.

We therefore show the expected share of volumes to be transferred between PMs during the transfer phase of BR under the different availability scenarios in Table 4.11. Overall, PMs are expected to trade 6.7% of the total volume of allocated AIs. By transferring 6.7% of AI volumes, the expected TAI can be reduced by 31.2%, from 173.8 for IO^U to 119.5 for BR.

Note that the findings from the analysis of the impact of shadow price approximation errors on expected performance are consistent with those obtained using synthetic data and are therefore omitted.

4.6.3. Distribution of shortage impact across PM

While our primary focus is on evaluating the performance of various allocation mechanisms with respect to the overall profit implications for the corporation, we also examine how the different allocation mechanisms distribute expected allocation impacts among PMs. This analysis is valuable for the case company for two reasons. First, it allows us to use the first-best solution of IO to determine the optimal level of differentiation. We then compare how well different allocation mechanisms replicate this optimal differentiation profile. Second, the analysis allows us to inform decision makers about the implications of implementing different mechanisms in practice beyond total expected profits. It thereby addresses past concerns regarding the degree of appropriate distribution of allocation burden across PMs.

We evaluate the degree to which the allocation impact is evenly distributed across the PMs by deriving the standard deviation of the ratio $r_f^{\mathcal{M}}$, which represents the expected profit impact of an individual PM under a given allocation mechanism relative to the expected profits if a PM receives the unconstrained, self-optimizing quantity:

$$r_f^{\mathcal{M}} = \frac{\pi(O_f^s) - \pi(Q_f^{\mathcal{M}})}{\pi(O_f^s)}. \tag{4.22}$$

In addition to the average standard deviation across the studied portfolios, we show the gap in percentage points (pp) between each method and the average standard deviation of IO.

		First-best	Proposal	Theoretical benchmark			In-company benchmark	
		IO	***BR***	***LA***	***IO***^{U}	***TP***	***SLA***	***UPA***
	Avg:	**9.0%**	**13.7% (4.7 pp)**	**21.7% (12.7 pp)**	**15.3% (6.4 pp)**	**14.6% (5.6 pp)**	**17.6% (8.6 pp)**	**3.3% (-5.6 pp)**
Avail	L/L	11.5%	17.2% (5.7 pp)	25.1% (13.7 pp)	18.5% (7.1 pp)	17.5% (6.0 pp)	20.5% (9.1 pp)	3.2% (-8.3 pp)
	L/H	15.4%	15.1% (-0.3 pp)	23.6% (8.3 pp)	18.1% (2.7 pp)	17.5% (2.1 pp)	19.4% (4.0 pp)	4.4% (-10.9 pp)
	H/L	8.3%	12.8% (4.5 pp)	22.2% (13.8 pp)	14.3% (6.0 pp)	12.9% (4.6 pp)	18.4% (10.0 pp)	4.7% (-3.7 pp)
	H/H	0.7%	9.7% (9.1 pp)	15.8% (15.2 pp)	10.5% (9.8 pp)	10.3% (9.7 pp)	12.1% (11.5 pp)	1.1% (0.5 pp)

Table 4.12.: Standard deviation of the share of unconstrained profits captured by PMs under different AI availability scenarios.

All methods except UPA lead to a higher degree of differentiation compared to IO. On average, BR most closely replicates the optimal degree of differentiation (+4.7 pp), followed by TP (+5.6 pp) and UPA (-5.6 pp). Both lexicographic methods lead to substantially higher levels of differentiation across PMs. By employing a two-step lexicographic approach, SLA reduces the gap from LA (+12.7 pp) to +8.6 pp. If the case company decides to implement either BR or UPA to replace SLA, it can improve overall performance while simultaneously distributing lost profits due to limited AI availability more evenly across its PMs.

4.7. Conclusion

In this paper, we addressed a real-world allocation problem faced by an agrochemical supplier of CPPs. At the case company, the production and distribution of products are handled centrally by the HQ, while distributed PMs are dedicated to selling a given product to local customers within a market territory. CPPs are produced by blending one or multiple AIs into final products that are applied by farmers worldwide. Prior to the start of the selling season, the HQ of the case company faces the allocation problem of determining how much of each final product to blend from the limited available AI volumes in order to maximize the firm's expected total profits. The allocation is complicated by the high uncertainty in the market demand for CPPs. Based on frequent interactions with local customers, only the local PM can estimate the probability of different demand scenarios in their market. This knowledge is private to the responsible PM and not known to the HQ or other PMs. To leverage this proprietary market knowledge, the HQ asks PMs to submit product orders prior to the selling season. As PMs are compensated based on local market profits, the allocation process is challenged by PMs leveraging this information asymmetry by strategically ordering products from the HQ in an attempt to maximize their own profitability rather than the firm's overall profitability.

To support the allocation problem of the case company, we developed a hybrid allocation mechanism that is applicable in the context of distributed decision making, private information, and conflicting incentives.

We assessed the performance of our proposed approach against current company practice and several benchmarks from the literature in a numerical study based on a controlled synthetic data set and real-world data from the case company. We find that our proposed method (BR) leads on average to expected lost profits that are 29.0% higher than those obtainable under the first-best solution under full central information availability for the studied real-world data of the case company. Our method thereby closes two-thirds of the gap between the current company practice (SLA) and the first-best solution (91.9%).

Leveraging the proposed approach, which combines a top-down allocation in the first step with an opportunity for PMs to adjust the initial allocation bottom-up in a second step in exchange for a transfer payment, results in better performance than using either step in isolation. Any bottom-up adjustment of an initial top-down allocation improves the firm's total expected profit, as self-optimizing PMs request additional or reduced volumes only when the expected impact on local profits exceeds the associated transfer payment owed or received. By reallocating AIs to PMs with a higher willingness to pay, the expected profits of the firm are improved. In addition, we find that the performance of our proposed mechanism is substantially more robust to misspecified transfer prices than a pure transfer-pricing approach.

Our work opens several opportunities for future research. First, the goal of this paper was to propose an allocation mechanism applicable within the existing incentive scheme of the case company, in which PMs are compensated based on local profits. Exploring how insights from the mechanism design literature could be applied to fundamentally redesign the incentive scheme to address intrafirm conflicts would be an interesting avenue for future work. Second, we discussed the allocation of limited AIs to final products as a single-period problem. In reality, PMs might consider expectations regarding availability scenarios in future selling seasons when engaging in strategic interactions with the HQ during the current season. Extending the analysis to a multi-season allocation problem would be worthwhile. Finally, we captured the interaction of two groups of agents through a bi-level game. In reality, more agents might strategically interact with each other. For the discussed allocation problem, it might be particularly interesting to analyze the interaction between the HQ, multiple regionally managed production sites, and local product managers.

Appendix A

Appendix to Chapter II: Solution heuristic and computational improvements

In the following section, we present a naïve solution heuristic to solve the budget planning model (Section A.1), for which we subsequently propose several computational improvements (Section A.2).

A.1. Naïve solution heuristic

We solve the discussed budget planning problem with a gradient search approach to determine the minimum expected lost margin volume guarantee, $\vec{v}^*$, which ensures that the associated inventory investment, assessed by the AARO nested in subproblem (2.5) - (2.6), complies with the maximum allowable inventory investment.

As outlined in Section 2.5.1 and equation (2.77), we follow the practice of the case company to simplify the derivation of volume guarantees by only determining overall guarantee levels by region and AI which are subsequently disaggregated to a monthly guarantee plan such that the same share of the uncertain demand component is covered along all months of the budget year:

$$v_{r,a,t}(g_{r,a}) = d^-_{r,a,t} + g_{r,a}(d^+_{r,a,t} - d^-_{r,a,t}) \quad \forall r \in \mathcal{R}, a \in \mathcal{A}, t \in \mathcal{T}. \tag{A.1}$$

Initially, we set all guarantees, $v_{r,a,t}$, such that they cover the full support of uncertain demand distributions: $g_{r,a} \leftarrow 1 \, \forall r \in \mathcal{R}, a \in \mathcal{A}$. Supplying maximum volume guarantees to all regions is most likely incompatible with the maximum allowable inventory investment β. We, therefore, iteratively select and subsequently reduce regional AI volume guarantees by a fixed step size ϵ until the associated total worst-case inventories no longer exceed β. To decide which guarantee to cut in each search iteration, the expected impact of reducing a single regional AI guarantee ($g_{r,a,}$) in a given guarantee level vector ($\vec{g}$) by ϵ is assessed based on the implied change in expected lost margin ($\delta LM(\vec{g}_{r,a})$) relative to the change in worst-case inventories ($\delta WC(\vec{g}_{r,a})$), given by the following gradient:

$$\gamma_{r,a} = \frac{\delta LM(\vec{g}_{r,a})}{\delta WC(\vec{g}_{r,a})}. \tag{A.2}$$

We determine the lost margin-minimal $\vec{g}^*$ by iteratively selecting and reducing the volume guarantee with the highest gradient value, i.e., the regional AI guarantee with the lowest expected lost margin increase per unit of worst-case inventory decrease. In each iteration, we reduce the chosen regional AI guarantee by step size ϵ, update the gradient values of all regional AI guarantees based on the latest $g_{r,a}$ and continue in this fashion as long as the associated worst-case inventories exceed the maximum allowable inventory investment. Note that due to the company's business rule that at least the certain component of uncertain regional AI demand must be supplied, $g_{r,a} \geq 0 \, \forall r \in \mathcal{R}, a \in \mathcal{A}$ is a lower bound. We summarize the search approach in Algorithm A.1.

Algorithm A.1: Naïve search

Result: $\vec{g}^*$
$g_{r,a} \leftarrow 1 \, \forall r \in \mathcal{R}, a \in \mathcal{A}$;
while $\beta < WC(\vec{g})$ **do**
  Update $\gamma_{r,a}$ based on $\vec{g} \quad \forall r \in \mathcal{R}, a \in \mathcal{A}$;
  $(r', a') \leftarrow argmax(\gamma_{r,a} : g_{r,a} - \epsilon \geq 0)$;
  $g_{r',a'} \leftarrow g_{r',a'} - \epsilon$;
$\vec{g}^* \leftarrow \vec{g} = (g_{1,1}, \ldots, g_{R,A})$

A.2. Computational improvements of local search

The proposed search procedure can be computationally intense even for medium-sized problems. In each search iteration all $r \times a$ gradients must be updated, each requiring us to solve an AARO problem instance in order to derive worst-case inventory deltas. Even though the AARO reformulation preserves the LP model structure of the original static RO, the overall problem size substantially increases as each static decision variable is transformed into an affine policy specified through a $(T+1)$-dimensional decision vector together with T additional auxiliary variables required for the reformulation of the robust constraints. Therefore, we strive to solve as few AARO model instances as possible. We propose two levers to improve the overall search performance: first, reducing the solution time of an individual AARO model through reducing the size of the coefficient matrices (Section A.2.1) and, second, limiting the number of gradients recalculated per search iteration (Section A.2.2).

A.2.1. Reducing the AARO model size

In the AARO formulation presented in Section 2.4, decisions in period t can be affinely conditioned on all realized uncertainties up to period $t - \delta$ through the respective coefficient matrices. In line with observations from Ben-Tal et al. (2005) and See and Sim (2010), the solution structure of the coefficient matrices obtained when solving the AARO is commonly sparse, i.e., decisions are adjusted based on only few and recently revealed uncertainties. Intuitively, as new actual sales realizations become available, supply chain will adjust affected production and ordering decisions, while considering inventories, remaining volume guarantees and relevant constraints such as lead times or capacities. After the decision adjustments in response to an actual monthly regional AI sale are completed, no further future decisions in subsequent periods are influenced by the previously revealed uncertainty.

Based on this observation, we introduce a look-back period (LBP) to reduce the dimensions of the coefficient matrices specifying the decision policies. We, thereby, reduce the overall size of the AARO model formulation by ensuring that decisions in period t can only be affinely adjusted based on uncertainty realizations of the last LBP periods. Thus, the dimensionality of a given coefficient matrix such as $O^{\tau}_{r,a,t}$ can be reduced from $R \times A \times T \times T$ to $R \times A \times T \times LBP$. Through the specification of LBP, we can trade off the complexity of affine policies, the model size and the solution time against the solution quality of a tight and not overly conservative worst-case inventory assessment. To derive the smallest possible LBP which does not lead to a deterioration of the solution quality for a given problem instance, i.e., an increase in the derived worst-case objective function value, we initially solve the AARO under an unrestricted information base (i.e., $LBP = T$) for maximum volume guarantees across all regions and AIs ($g_{r,a} = 1 \; \forall r \in \mathcal{R}, a \in \mathcal{A}$). Subsequently, we gradually reduce the LBP by one period as long as the objective value obtained when solving the restricted AARO remains unchanged and the solution quality of the worst-case inventory assessment is thus not yet affected by the reduced size of the coefficient matrices specifying the decision policies. The process is depicted in Algorithm A.2.

Algorithm A.2: Determining look-back period (LBP)

Result: LBP^*
$g_{r,a} \leftarrow 1 \quad \forall r \in \mathcal{R}, a \in \mathcal{A}$;
$wc \leftarrow WC_T(\vec{g})$; `/* Solution from original AARO with full LBP T */`
$n \leftarrow T - 1$;
while $WC_n(\vec{g}) = wc \,\&\, n \geq 0$ **do**
  $n \leftarrow n - 1$;
$LBP^* \leftarrow n$;

Note that, through Algorithm A.2, we conservatively determine the LBP for a problem instance with volume guarantees corresponding to the upper support of uncertain demand in order to maximize the need for worst-case decision adjustments in response to AI sales under low demands. As volume guarantees decrease along the gradient search process and the gap between volume guarantees and worst-case sales under low demands is reduced, the number of periods required to adjust supply chain decisions in response to worst-case sales realizations by, for example, cutting production and regional orders, is non-increasing. Eventually, for the case of $g_{r,a} \leq 0 \, \forall r \in \mathcal{R}, a \in \mathcal{A}$, the LPB will be 0 as guaranteed volumes will always be retrieved in full to meet certain demands and no decisions must be adjusted.

In addition, the derivation of the LBP offers an interesting perspective on the reactiveness of the supply chain in dealing with demand-side signals as LBP can be interpreted as the maximum number of periods required to actively react to realized uncertainties through decision adjustments until any impact on worst-case performance is fully mitigated.

By introducing the LBP, we can cut the average number of constraints by 58%, the number of decision variables by 59%, and the computation time per AARO model run by 93% from about 60 seconds to 4 seconds for the synthetic data set.[1]

A.2.2. Reducing the number of gradients to be updated

After improving the computation time per gradient, we focus on how to limit the number of gradients to be updated per search iteration. We therefore examine if and when a gradient $\gamma^{i}_{r',a'}$ updated in iteration i can be utilized in any following search iteration j $(i < j)$.

In the gradient's numerator, the delta in expected lost margin $(\delta LM_{r',a'}(\vec{g}_{r',a'}))$ when reducing the provided regional AI guarantee $g_{r',a'}$ by step size ϵ is independent of the AI volumes provided to other regions. Any marginal expected lost margin assessment derived in iteration i for (r', a') is thus unaffected by potential reductions in other regional AI guarantees and remains valid in subsequent iterations until the regional AI guarantee itself is selected for reduction.

For assessing worst-case inventories, the supply plan of AI a' to its selling regions is, in the proposed model formulation, fully decoupled from decisions related to other AIs due to both dedicated global production capacities per AI and separated regional handing. For any (r', a'), a worst-case inventory delta as part of the gradient's denominator assessed in iteration i remains therefore valid for any succeeding iteration j $(i < j)$ as long as no regional guarantee of the same AI a' has been cut in the meantime.

By utilizing both observations, we can reduce the number of gradients to be updated per iteration from $r \times a$ to only r by solely updating the gradients of all regions associated with the AI selected for reduction during the iteration ($\gamma_{r,a'}\ \forall r \in 1, ..., R$). For the synthetic data set with 18 AIs supplied to three regions, this corresponds in a reduction of 94% in the number of gradients to be updated per search iteration.

[1] Executed on an Intel(R) Core(TM) i5-8350U CPU with 16 GB memory using FICO Xpress Optimizer Version 30.01.04.

Both, the reduced information base (LBP) and the limited gradient updates, thus substantially enhance the computational efficiency of the basic gradient search introduced in Section A.1. The refined search procedure is outlined in Algorithm A.3.

Algorithm A.3: Computationally improved search

```
Result: g⃗*
g_{r,a} ← 1 ∀r ∈ R, a ∈ A;
n ← LBP(g⃗);
for r ∈ R, a ∈ A do
    Calculate initial γ_{r,a} based on g⃗ ∀r ∈ R, a ∈ A;
while β < WC(g⃗) do
    (r', a') ← argmax(γ_{r,a} : g_{r,a} − ε ≥ 0);
    g_{r',a'} ← g_{r',a'} − ε;
    Update γ_{r,a'} based on updated g⃗ ∀r ∈ R
g⃗* ← g⃗ = (g_{1,1}, . . . , g_{R,A})
```

Appendix B

Appendix to Chapter III: Proofs

In the following section, we present the proofs of Proposition 3.1 to Proposition 3.3.

B.1. Proof of Proposition 3.1

By definition of the decision problem, Q_i and $R_i(d_i)$ must be chosen for a given τ_i such that the total actual inventory investment (l_i) of the BU does not exceed the BU-specific inventory investment ceiling (τ_i) under any realization of d_i:

$$l_i \leq \tau_i \quad \forall d_i \in \mathcal{D}_i. \tag{B.1}$$

Since we assume that per unit inventory reduction cost exceeds the capital cost of holding one unit of leftover product until the end of the business year, i.e., $k_i > v_i \cdot h_i \cdot t_i^a$, leftover inventories will only be reduced if it is necessary to ensure adherence to the imposed inventory investment constraint.

To ensure that reducing leftover inventories is sufficient to ensure robust adherence to the total inventory investment ceiling under any realization of $d_i \in \mathcal{D}_i$, the deterministic pre-production inventories associated with the supply of Q_i cannot exceed the total inventory investment ceiling, i.e., $\tau_i \geq v_i \cdot (t_i^b \cdot s_i + \frac{(Q_i - s_i)^2}{2 \cdot c_i})$.

If $\tau_i \geq v_i \cdot (t_i^b \cdot s_i + \frac{(Q_i - s_i)^2}{2 \cdot c_i})$, we can therefore reformulate the total inventory investment of BU i expressed in equation (3.1) to

$$l_i = min\{\tau_i, v_i \cdot (t_i^b \cdot s_i + \frac{(Q_i - s_i)^2}{2 \cdot c_i} + t_i^a \cdot (Q_i - d_i)^+)\}. \tag{B.2}$$

For a given Q_i and τ_i, λ_i represents the lowest demand realization at which all resulting leftovers $(Q_i - \lambda_i)$ can be kept in stock without exceeding the total inventory investment ceiling τ_i. λ_i can thus be interpreted as the inventory reduction threshold, i.e., any additional leftover volumes caused by demand realizations below λ_i must be fully reduced, leading to the optimal reduction decision expressed in equation (3.11).

B.2. Proof of Proposition 3.2

The optimal supply quantity Q_i^* must be feasible with respect to constraint (3.13) and constraint (3.17). We rearrange constraint (3.17) to derive the maximum supply quantity $\overline{Q}_i$ that can be supplied under τ_i as defined in expression (3.19).

We thus know that

$$s_i \leq Q_i^* \leq \overline{Q}_i \tag{B.3}$$

and that the problem is infeasible if available permits are not sufficient to cover the pre-season inventory investment associated with the supply of starting inventories s_i, i.e., $v_i \cdot t_i^b \cdot s_i \leq \tau_i$.

We take the first derivative of the objective function

$$\begin{aligned} \frac{\partial}{\partial Q_i} E[\pi_i] = m_i(1 - F_i(Q_i)) - v_i \cdot h_i(\frac{Q_i - s_i}{c_i} + t_i^a \cdot F_i(Q_i)) \\ - (1 + \frac{Q_i - s_i}{c_i \cdot t_i^a}) \cdot (k_i - v_i \cdot h_i \cdot t_i^a) \cdot F_i(\lambda_i) = 0 \end{aligned} \tag{B.4}$$

to solve for the interior solution (Q_i^{int}) as denoted in expression (3.20).

The optimization problem is convex, as the objective function to be maximized is concave

$$\begin{aligned} (\frac{\partial}{\partial Q_i})^2 E[\pi_i] = &- (m_i + v_i \cdot h_i \cdot t_{i^a}) \cdot f_i(Q_i) - \frac{v_i \cdot h_i}{c_i} \\ &- (1 + \frac{Q_i - s_i}{c_i \cdot t_i^a})^2 \cdot (k_i - v_i \cdot h_i \cdot t_i^a) \cdot f_i(\lambda_i) - \frac{k_i - v_i \cdot h_i \cdot t_i^a}{c_i \cdot t_i^a} \cdot F_i(\lambda_i) \leq 0 \end{aligned} \tag{B.5}$$

as all parameters are non-negative, $k_i > v_i \cdot h_i \cdot t_i^a$ and both constraints are convex.

The interior solution is thus optimal if Q_i^{int} violates neither of the constraints, i.e., $s_i \leq Q_i^{int} \leq \overline{Q}_i$. Otherwise, Q_i^* is characterized by either of the two bounds, i.e., $\overline{Q}_i$ if $s_i \leq \overline{Q}_i < Q_i^{int}$ or s_i if $Q_i^{int} < s_i \leq \overline{Q}_i$.

B.3. Proof of Proposition 3.3

To derive optimal decision policies, we define μ_i as the Lagrange multiplier of constraint (3.25) and γ_i as the Lagrange multiplier of constraint (3.26). The Lagrangian function is:

$$L(Q_i, \tau_i, \mu_i, \gamma_i) = E[\pi_i] - p \cdot \tau_i + \mu_i(\tau_i - v_i(t_i^b \cdot s_i + \frac{(Q_i - s_i)^2}{2 \cdot c_i})) + \gamma_i(Q_i - s_i) \quad \text{(B.6)}$$

We obtain the following first-order Karush-Kuhn-Tucker (KKT) conditions:

Stationarity conditions:

$$\begin{aligned}
\frac{\partial}{\partial Q_i} L(Q_i, \tau_i, \mu_i, \gamma_i) =& m_i(1 - F_i(Q_i)) - v_i \cdot h_i(\frac{Q_i - s_i}{c_i} + t_i^a \cdot F_i(Q_i)) \qquad \text{(B.7)} \\
& - (1 + \frac{Q_i - s_i}{c_i \cdot t_i^a}) \cdot (k_i - v_i \cdot h_i \cdot t_i^a) \cdot F_i(\lambda_i) \\
& - \mu_i(\frac{v_i(Q_i - s_i)}{c_i}) + \gamma_i = 0 \\
\frac{\partial}{\partial \tau_i} L(Q_i, \tau_i, \mu_i, \gamma_i) =& (\frac{k_i}{v_i \cdot t_i^a} - h_i) \cdot F_i(\lambda_i) - p + \mu_i = 0 \qquad \text{(B.8)}
\end{aligned}$$

Primal feasibility:

$$v_i(t_i^b \cdot s_i + \frac{(Q_i - s_i)^2}{2 \cdot c_i})) - \tau_i \leq 0 \qquad \text{(B.9)}$$

$$s_i - Q_i \leq 0 \qquad \text{(B.10)}$$

Complementary slackness:

$$\mu_i(v_i(t_i^b \cdot s_i + \frac{(Q_i - s_i)^2}{2 \cdot c_i})) - \tau_i) = 0 \qquad \text{(B.11)}$$

$$\gamma_i(Q_i - s_i) = 0 \qquad \text{(B.12)}$$

Dual feasibility:

$$\mu_i \geq 0 \tag{B.13}$$

$$\gamma_i \geq 0 \tag{B.14}$$

Convexity of optimization problem:

The KKT conditions are necessary and sufficient if the optimization problem is convex. For this to hold, the objective function must be concave in Q_i and τ_i. We take the following second-order derivatives:

$$\frac{\partial E[\pi_i]^2}{\partial Q_i^2} = -(m_i + v_i \cdot h_i \cdot t_{i^a}) \cdot f_i(Q_i) - \frac{v_i \cdot h_i}{c_i} \tag{B.15}$$

$$- (k_i - v_i \cdot h_i \cdot t_i^a) \cdot (\frac{F_i(\lambda_i)}{c_i \cdot t_i^a} + (1 + \frac{Q_i - s_i}{c_i \cdot t_i^a})^2 \cdot f_i(\lambda_i)) \tag{B.16}$$

$$\frac{\partial E[\pi_i]^2}{\partial \tau_i^2} = -\frac{1}{v_i \cdot t_i^a}(\frac{k_i}{v_i \cdot t_i^a} - h_i) f_i(\lambda_i) \tag{B.17}$$

$$\frac{\partial E[\pi_i]^2}{\partial Q_i \partial \tau_i} = (1 + \frac{Q_i - s_i}{c_i \cdot t_i^a})(\frac{k_i}{v_i \cdot t_i^a} - h_i) f_i(\lambda_i) \tag{B.18}$$

The objective function is concave if the following conditions apply:

$$\frac{\partial E[\pi_i]^2}{\partial Q_i^2} \leq 0 \tag{B.19}$$

$$\frac{\partial E[\pi_i]^2}{\partial \tau_i^2} \leq 0 \tag{B.20}$$

$$\frac{\partial E[\pi_i]^2}{\partial Q_i^2} \cdot \frac{\partial E[\pi_i]^2}{\partial \tau_i^2} - (\frac{\partial E[\pi_i]^2}{\partial Q_i \partial \tau_i})^2 \geq 0 \tag{B.21}$$

Condition (B.19) and condition (B.20) are fulfilled as $k_i > v_i \cdot h_i \cdot t_i^a$. We can plug the second order derivatives into condition (B.21) and rearrange to

$$((\frac{m_i}{v_i \cdot t_i^a} + h_i) \cdot f_i(Q_i) + \frac{h_i}{c_i \cdot t_i^a}) \cdot (\frac{F_i(\lambda_i)}{c_i \cdot t_i^a} + (1 + \frac{Q_i - s_i}{c_i \cdot t_i^a})^2 \cdot f_i(\lambda_i)) + (\frac{k_i}{v_i \cdot t_i^a} - h_i) \cdot \frac{F_i(\lambda_i)}{c_i \cdot t_i^a} \geq 0$$

which holds as all parameters are non-negative and $k_i > v_i \cdot h_i \cdot t_i^a$. The KKT conditions are thus necessary and sufficient to characterize the optimal Q_i and τ_i.

We assess four cases based on whether constraints (3.25) and (3.26) are binding:

Case	Constraint (3.25)	Constraint (3.26)
1	not binding	not binding
2	not binding	binding
3	binding	not binding
4	binding	binding

Table B.1.: Cases to be assessed via KKTs.

Case 1: Interior solution ($\mu_i = 0, \gamma_i = 0$)

By rearranging the stationary condition (B.8), we obtain $\lambda_i^{C1} = F_i^{-1}(\frac{p \cdot v_i \cdot t_i^a}{k_i - v_i \cdot h_i \cdot t_i^a})$ which we insert into the stationary condition (B.7) to obtain the optimality condition for Q_i^{C1}

$$Q_i^{C1} : m_i + \frac{h_i + p}{c_i} \cdot v_i \cdot s_i = p \cdot v_i \cdot t_i^a + (m_i + v_i \cdot h_i \cdot t_i^a) \cdot F_i(Q_i^{C1}) + \frac{h_i + p}{c_i} \cdot v_i \cdot Q_i^{C1}. \tag{B.22}$$

To simplify the analysis in the following, we define the left-hand side of the optimality condition (B.22) as $A \equiv m_i + \frac{h_i+p}{c_i} \cdot v_i \cdot s_i$ and the right-hand side as $B(x) \equiv p \cdot v_i \cdot t_i^a + (m_i + v_i \cdot h_i \cdot t_i^a) \cdot F_i(x) + \frac{h_i+p}{c_i} \cdot v_i \cdot x$. Condition (B.22) is therefore equivalent to $Q_i^{C1} : A = B(Q_i^{C1})$, which we can solve numerically.

We rearrange the definition of λ_i in equation (3.12) to obtain the optimal number of purchased permits τ_i for a known Q_i and λ_i: $\tau_i^{C1} = v_i \cdot (t_i^b \cdot s_i + \frac{(Q_i^{C1} - s_i)^2}{2 \cdot c_i} + t_i^a \cdot (Q_i^{C1} - \lambda_i^{C1}))$.

For Q_i^{C1}, λ_i^{C1}, and τ_i^{C1} to be the optimal solution to the problem, it must satisfy primal feasibility of constraint (B.9) and constraint (B.10). Constraint (B.9) is satisfied whenever $Q_i^{C1} \geq \lambda_i^{C1}$, i.e., the number of purchased permits covers at least the deterministic pre-season inventory investment. As $B(x)$ is non-decreasing in x and $A = B(Q_i^{C1})$, we know that constraint (B.9) is fulfilled if $A \geq B(\lambda_i^{C1} = F_i^{-1}(\frac{p \cdot v_i \cdot t_i^a}{k_i - v_i \cdot h_i \cdot t_i^a}))$. Similarly, constraint (B.10) is ensured if $A \geq B(s_i)$ which we can rearrange to $\frac{m_i - p \cdot v_i \cdot t_i^a}{m_i + v_i \cdot h_i \cdot t_i^a} \geq F_i(s_i)$.

Case 2: $\mu_i = 0, Q_i^{C2} = s_i$

By definition of case 2, $Q_i^{C2} = s_i$. Based on the stationary condition (B.8) and as $\mu_i = 0$, we know that $\lambda_i^{C2} = F_i^{-1}(\frac{p \cdot v_i \cdot t_i^a}{k_i - v_i \cdot h_i \cdot t_i^a})$ and $\tau_i^{C2} = v_i \cdot (t_i^b \cdot s_i + t_i^a \cdot (s_i - F_i^{-1}(\frac{p \cdot v_i \cdot t_i^a}{k_i - v_i \cdot h_i \cdot t_i^a})))$. Case 2 is the applicable if Q_i^{C2}, λ_i^{C2} and τ_i^{C2} fulfill primal and dual feasibility.

The stationary condition requires $Q_i^{C2} : A = B(Q_i^{C2}) - \gamma_i$ and, as $Q_i^{C2} = s_i$, $A + \gamma_i = B(s_i)$. Primal feasibility of constraint (B.9) is ensured if $Q_i^{C2} = s_i \geq \lambda_i^{C2}$ or equivalently $A + \gamma_i = B(s_i) \geq B(F_i^{-1}(\frac{p \cdot v_i \cdot t_i^a}{k_i - v_i \cdot h_i \cdot t_i^a}))$.

We can check for dual feasibility ($\gamma_i \geq 0$) by rearranging the stationarity condition (B.7) for γ_i

$$\gamma_i = -m_i(1 - F_i(s_i)) + v_i \cdot h_i \cdot t_i^a \cdot F_i(s_i) + p \cdot v_i \cdot t_i^a. \tag{B.23}$$

which will be non-negative if $\frac{m_i - p \cdot v_i \cdot t_i^a}{m_i + v_i \cdot h_i \cdot t_i^a} \leq F_i(s_i)$.

Case 3: $\tau_i^{C3} = v_i(t_i^b \cdot s_i + \frac{(Q_i^{C3} - s_i)^2}{2 \cdot c_i}), \gamma_i = 0$

Under case 3, constraint (3.25), ensuring that at least the certain pre-season inventory investment is covered by inventory investment permits purchased during the auction, is binding. Consequently, any leftovers must be fully reduced after the selling season to adhere to the inventory investment ceiling, i.e., $Q_i^{C3} = \lambda_i^{C3}$. We set $Q_i^{C3} = \lambda_i^{C3}$, rearrange the stationary condition (B.8) for μ_i, and insert the term into stationary condition (B.7). We obtain the following optimality condition for Q_i^{C3} and λ_i^{C3}:

$$Q_i^{C3} : m_i + (p + h_i) \cdot \frac{v_i \cdot s_i}{c_i} = (m_i + k_i) \cdot F_i(Q_i^{C3}) + (p + h_i) \cdot \frac{v_i \cdot Q_i^{C3}}{c_i} \tag{B.24}$$

In line with case 1, we define $C(x)$ as the right-hand side of equation (B.24), i.e., $C(x) \equiv (m_i + k_i) \cdot F_i(x) + (p + h_i) \cdot \frac{v_i \cdot x}{c_i}$.

Condition (B.24) is thus equivalent to $Q_i^{C3} : A = C(Q_i^{C3})$ which we can solve numerically. Optimal decisions under case 3 are: $Q_i^{C3} : A = C(Q_i^{C3})$, $\lambda_i^{C3} = Q_i^{C3}$, $\tau_i^{C3} = v_i \cdot (t_i^b \cdot s_i + \frac{(Q_i^{C3} - s_i)^2}{2 \cdot c_i})$.

Primal feasibility ($s_i \leq Q_i^{C3}$) is fulfilled if $A = C(Q_i^{C3}) \geq C(s_i)$ which we can rearrange to $\frac{m_i}{m_i + k_i} \geq F_i(s_i)$.

Based on condition (B.8), dual feasibility ($\mu_i \geq 0$) requires that $\mu_i = -(\frac{k_i}{v_i \cdot t_i^a} - h_i) \cdot F_i(\lambda_i^{C3}) + p \geq 0$, which is equivalent to $F_i^{-1}(\frac{p \cdot v_i \cdot t_i^a}{k_i - v_i \cdot h_i \cdot t_i^a}) \geq \lambda_i^{C3}$. For λ_i^{C3} and Q_i^{C3} to be the optimal solution, it must thus not exceed the λ_i^{C1} of the interior solution (case 1) which is ensured if $A \leq C(F_i^{-1}(\frac{p \cdot v_i \cdot t_i^a}{k_i - v_i \cdot h_i \cdot t_i^a}))$. As $B(F_i^{-1}(\frac{p \cdot v_i \cdot t_i^a}{k_i - v_i \cdot h_i \cdot t_i^a})) = C(F_i^{-1}(\frac{p \cdot v_i \cdot t_i^a}{k_i - v_i \cdot h_i \cdot t_i^a}))$, this is equivalent to $A \leq B(F_i^{-1}(\frac{p \cdot v_i \cdot t_i^a}{k_i - v_i \cdot h_i \cdot t_i^a}))$.

Case 4: $Q_i^{C4} = \lambda_i^{C4} = s_i$, $\tau_i^{C4} = t_i^b \cdot s_i \cdot v_i$

In case 4, both constraints are binding. Complementary slackness and primary feasibility requires that $Q_i^{C4} = s_i$, $\lambda_i^{C4} = s_i$, $\tau_i^{C4} = t_i^b \cdot s_i \cdot v_i$. We rearrange the stationary condition $Q_i^{C4} : A = C(Q_i^{C4}) - \gamma_i$ to check for $\gamma_i \geq 0$ and obtain $\gamma_i = -m_i(1 - F_i(s_i)) + k_i \cdot F_i(s_i) \geq 0$. We can simplify to $F_i(s_i) \geq \frac{m_i}{m_i + k_i}$.

Similarly, we rearrange the stationary condition (B.8) and check for $\mu_i \geq 0$. The condition holds if $\mu_i = p - F_i(s_i)(\frac{k_i}{v_i \cdot t_i^a} - h_i) \geq 0$, which we simplify to $\frac{p \cdot v_i \cdot t_i^a}{k_i - h_i \cdot v_i \cdot t_i^a} \geq F_i(s_i)$. To simplify the split of optimal decision policies, this can equivalently be expressed as $B(F_i^{-1}(\frac{p \cdot v_i \cdot t_i^a}{k_i - v_i \cdot h_i \cdot t_i^a})) \geq B(s_i)$.

Note that $B(F_i^{-1}(\frac{p \cdot v_i \cdot t_i^a}{k_i - v_i \cdot h_i \cdot t_i^a})) = C(F_i^{-1}(\frac{p \cdot v_i \cdot t_i^a}{k_i - v_i \cdot h_i \cdot t_i^a}))$ which we use to separate case 1 and 2 from case 3 and 4 in Proposition 3.3.

Appendix C

Appendix to Chapter IV: Algorithm of base-and-refine

Algorithm C.1 formalizes the full allocation process of the base-and-refine method presented in Section 4.4.3.

Algorithm C.1: Proposed allocation mechanism: Base-and-refine (BR).

Result: $\vec{Q}$

Order set of final products $f \in \mathcal{F} = \{1, ..., F\}$ such that $\frac{\sum_{a \in \mathcal{A}} i_{a,1}}{m_1} \leq ... \leq \frac{\sum_{a \in \mathcal{A}} i_{a,F}}{m_f}$;

Each PM f submits order (O_f) to HQ $(\vec{O} = (O_1, \ldots, O_F))$;

HQ derives base allocation ($\vec{B}$) and transfer prices ($\vec{\tau}$) ;

$\vec{B} \leftarrow IO^U$;

$\vec{\tau} \leftarrow \vec{\lambda}^{IO^U}$;

PMs determine transfer requests ($\vec{S}$) based on base allocation and transfer prices ;

for $f \leftarrow 1$ **to** F **do**

  $S_f = F^{-1}((\frac{m_f - \sum_{a \in \mathcal{A}} \tau_a \cdot i_{a,f}}{m_f + h_f})^+) - B_f$;

Continue on next page.

HQ matches transfer requests of PMs and confirms supply quantity $\vec{Q}$;
$\vec{Q} \leftarrow \vec{B}$;
$\vec{SAI} = (SAI_{1,1}, ..., SAI_{A,F})$; /* Transfer requests in pure AI volumes */
$\vec{SAI} \leftarrow (i_{1,1} \cdot S_1, ..., i_{A,F} \cdot S_F))$;
$\vec{TrAI} = (TrAI_{1,1,1}, ..., TrAI_{a,i,j}, ..., SAI_{A,F,F})$; /* Result matrix: Volumes of AI a shifted to product i from product j */
$\vec{TrAI} = (0, ..., 0)$;
for $f \leftarrow 1$ **to** $F-1$ **do**
  if $S_f > 0$ **then**
    $s \leftarrow min\{S_f, \underset{a \in A}{min}\{\sum_{j=1,\, i_{a,f}>0}^{F}((\frac{-SAI_{a,j}}{i_{a,j}})^{+})\}$; /* Determine how much of S_f can be fulfilled (s) based on opposing requests from other PMs */
    for $a \leftarrow 1$ **to** A *&* $i_{a,f} \cdot s > 0$ **do**
      $sai \leftarrow s \cdot i_{a,f}$; /* Determine volume of AI a to be swapped */
      $SAI_{a,f} \leftarrow SAI_{a,f} - sai$;
      Match PMs with opposing requests ;
      for $j \leftarrow 1$ **to** F *&* $sai > 0$ *&* $SAI_{a,j} < 0$ **do**
        $TrAI_{f,j,a} \leftarrow min\{sai, -SAI_{a,j}\}$;
        $SAI_{a,j} \leftarrow SAI_{a,j} + TrAI_{f,j,a}$;
        $sai \leftarrow sai - TrAI_{f,j,a}$;
  if $S_f < 0$ **then**
    $s \leftarrow max\{S_f, \underset{a \in A}{max}\{\sum_{j=1,\, i_{a,f}>0}^{F}((\frac{-SAI_{a,j}}{i_{a,j}})^{-})\}$; /* Determine how much of S_f can be fulfilled (s) based on opposing requests from other PMs */
    for $a \leftarrow 1$ **to** A *&* $i_{a,f} \cdot s > 0$ **do**
      $sai \leftarrow s \cdot i_{a,f}$; /* Determine volume of AI a to be swapped */
      $SAI_{a,f} \leftarrow SAI_{a,f} + sai$;
      Match PMs with opposing requests ;
      for $j \leftarrow 1$ **to** F *&* $sai < 0$ *&* $SAI_{a,j} > 0$ **do**
        $TrAI_{f,j,a} \leftarrow max\{sai, -SAI_{a,j}\}$;
        $SAI_{a,j} \leftarrow SAI_{a,j} + TrAI_{f,j,a}$;
        $sai \leftarrow sai - TrAI_{f,j,a}$;
for $f \leftarrow 1$ **to** F **do**
  $Q_f \leftarrow Q_f + \underset{a \in A,\, i_{a,f}>0}{min}\{\sum_{j=1}^{F} \frac{TrAI_{f,j,a}}{i_{a,f}}\}$; /* Determine final allocation quantity based on transferred AI volumes */

Bibliography

Affonso, R., Marcotte, F., and Grabot, B. (2008). Sales and operations planning: The supply chain pillar. *Production Planning and Control*, 19(2):132–141.

Agovino, M., Casaccia, M., Ciommi, M., Ferrara, M., and Marchesano, K. (2019). Agriculture, climate change and sustainability: The case of eu-28. *Ecological Indicators*, 105(April):525–543.

Agrawal, D. and Osadchiy, N. (2024). Inventory productivity and stock returns in manufacturing networks. *Manufacturing & Service Operations Management*, 26(2):573–593.

Agrawal, N. and Cohen, M. A. (2001). Optimal material control in an assembly system with component commonality. *Naval Research Logistics (NRL)*, 48(5):409–429.

Aharon, B.-T., Boaz, G., and Shimrit, S. (2009). Robust multi-echelon multi-period inventory control. *European Journal of Operational Research*, 199(3):922–935.

Alonso, R., Dessein, W., and Matouschek, N. (2008). When does coordination require centralization? *American Economic Review*, 98(1):145–179.

Ausubel, L. M. (2004). An efficient ascending-bid auction for multiple objects. *American Economic Review*, 94(5):1452–1475.

Ausubel, L. M., Cramton, P., Pycia, M., Rostek, M., and Weretka, M. (2014). Demand reduction and inefficiency in multi-unit auctions. *The Review of Economic Studies*, 81(4):1366–1400.

Aydin, N. and Birbil, S. I. (2018). Decomposition methods for dynamic room allocation in hotel revenue management. *European journal of operational research*, 271(1):179–192.

Bajgiran, O. S., Zanjani, M. K., and Nourelfath, M. (2016). The value of integrated tactical planning optimization in the lumber supply chain. *International Journal of Production Economics*, 171:22–33.

Bansal, S. and Dyer, J. S. (2020). Planning for end-user substitution in agribusiness. *Operations Research*, 68(4):1000–1019.

Bansal, S. and Nagarajan, M. (2017). Product portfolio management with production flexibility in agribusiness. *Operations Research*, 65(4):914–930.

Barahona, F., Bermon, S., Günlük, O., and Hood, S. (2005). Robust capacity planning in semiconductor manufacturing. *Naval Research Logistics (NRL)*, 52(5):459–468.

Bassett, M. and Gardner, L. (2010). Optimizing the design of global supply chains at dow agrosciences. *Computers & Chemical Engineering*, 34:254–265.

Beckmann, M. (1952). A lagrangian multiplier rule in linear activity analysis and some of its applications. *Cowles Commission Paper in Economics*, 2054.

Ben-Tal, A., Ghaoui, L. E., and Nemirovski, A. (2009). *Robust Optimization (Princeton Series in Applied Mathematics)*. Princeton University Press.

Ben-Tal, A., Golany, B., Nemirovski, A., and philippe Vial, J. (2005). Retailer-supplier flexible commitments contracts: A robust optimization approach. *Manufacturing and Service Operations Management*, 7(3):248–271.

Ben-Tal, A., Goryashko, A., Guslitzer, E., and Nemirovski, A. (2004). Adjustable robust solutions of uncertain linear programs. *Mathematical Programming*, 99(2):351–376.

Benjaafar, S., Chen, D., and Wang, R. (2017). Managing production-inventory systems with scarce resources. *Manufacturing & Service Operations Management*, 19(2):216–229.

Bertsimas, D., Brown, D. B., and Caramanis, C. (2011). Theory and applications of robust optimization. *SIAM Review*, 53(3):464–501.

Bhagwat, R. and Sharma, M. K. (2007). Performance measurement of supply chain management: A balanced scorecard approach. *Computers & Industrial Engineering*, 53(1):43–62.

Birge, J. R. and Louveaux, F. (2011). *Introduction to stochastic programming*. Springer Science & Business Media.

Bretthauer, K. M. and Shetty, B. (1995). The nonlinear resource allocation problem. *Operations research*, 43(4):670–683.

Bretthauer, K. M., Shetty, B., Syam, S., and Vokurka, R. J. (2006). Production and inventory management under multiple resource constraints. *Mathematical and Computer Modelling*, 44(1-2):85–95.

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

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

Cachon, G. P. and Lariviere, M. A. (1999c). An equilibrium analysis of linear, proportional and uniform allocation of scarce capacity. *Iie Transactions*, 31(9):835–849.

Celikbas, M., Shanthikumar, J. G., and Swaminathan, J. M. (1999). Coordinating production quantities and demand forecasts through penalty schemes. *IIE Transactions (Institute of Industrial Engineers)*, 31(9):851–864.

Chen, F., Li, J., and Zhang, H. (2013). Managing downstream competition via capacity allocation. *Production and Operations Management*, 22(2):426–446.

Chen, T.-L. and Lu, H.-C. (2012). Stochastic multi-site capacity planning of tft-lcd manufacturing using expected shadow-price based decomposition. *Applied Mathematical Modelling*, 36(12):5901–5919.

Chen, X. (2007). A robust optimization perspective on stochastic programming. *Operations Reserach*, 55(6):1058–1071.

Chen, X., Sim, M., Sun, P., and Zhang, J. (2008). A linear decision-based approximation approach to stochastic programming. *Operations Research*, 56(2):344–357.

Claro, J. and De Sousa, J. P. (2012). A multiobjective metaheuristic for a mean-risk multistage capacity investment problem with process flexibility. *Computers & Operations Research*, 39(4):838–849.

Comhaire, P. and Papier, F. (2015). Syngenta uses a cover optimizer to determine production volumes for its european seed supply chain. *Interfaces*, 45(6):501–513.

Cooper, R. G., Edgett, S. J., and Kleinschmidt, E. J. (2000). New problems, new solutions: making portfolio management more effective. *Research-Technology Management*, 43(2):18–33.

Cramton, P. and Kerr, S. (2002). Tradeable carbon permit auctions how and why to auction not grandfather. *Energy Policy*, 30(4):333–345.

Curry, R. E. (1990). Optimal airline seat allocation with fare classes nested by origins and destinations. *Transportation science*, 24(3):193–204.

de Farias, D. P. and Roy, B. V. (2003). The linear programming approach to approximate dynamic programming. *Operations Research*, 51(6):850–865.

Erlebacher, S. J. (2000). Optimal and heuristic solutions for the multi-item newsvendor problem with a single capacity constraint. *Production and Operations Management*, 9(3):303–318.

Ertogral, K. and Wu, S. D. (2000). Auction-theoretic coordination of production planning in the supply chain. *IIE Transactions*, 32(10):931–940.

Feng, Y., D'Amours, S., and Beauregard, R. (2008). The value of sales and operations planning in oriented strand board industry with make-to-order manufacturing system: Cross functional integration under deterministic demand and spot market recourse. *International Journal of Production Economics*, 115(1):189–209.

Feng, Y., Martel, A., D'Amours, S., and Beauregard, R. (2013). Coordinated contract decisions in a make-to-order manufacturing supply chain: A stochastic programming approach. *Production and Operations Management*, 22(3):642–660.

Fritz, M. and Hausen, T. (2009). Electronic supply network coordination in agrifood networks: Barriers, potentials, and path dependencies. *International Journal of Production Economics*, 121(2):441–453.

Gehen, S., Corvaro, M., Jones, J., Ma, M., and Yang, Q. (2019). Challenges and opportunities in the global regulation of crop protection products. *Organic Process Research & Development*, 23(10):2225–2233.

Goh, S. H. and Eldridge, S. (2019). Sales and operations planning: The effect of coordination mechanisms on supply chain performance. *International Journal of Production Economics*, 214(May 2018):80–94.

Gorissen, B. L., Yanikoglu, I., and den Hertog, D. (2015). A practical guide to robust optimization. *Omega*, 53:124–137.

Grimson, J. A. and Pyke, D. F. (2007). Sales and operations planning: an exploratory study and framework. *The International Journal of Logistics Management*, 18(3):322–346.

Groves, T. and Loeb, M. (1979). Incentives in a divisionalized firm. *Management Science*, 25(3):221–230.

Gunasekaran, A., Patel, C., and Tirtiroglu, E. (2001). Performance measures and metrics in a supply chain environment. *International Journal of Operations & Production Management*, 21(1/2):71–87.

Hadley, G. and Within, T. (1963). Analysis of inventory systems. Prentice-Hall.

Hahn, G. J. and Kuhn, H. (2011). Optimising a value-based performance indicator in mid-term sales and operations planning. *Journal of the Operational Research Society*, 62(3):515–525.

Harris, M., Kriebel, C. H., and Raviv, A. (1982). Asymmetric information, incentives and intrafirm resource allocation. *Management Science*, 28(6):604–620.

Hausen, T., Fritz, M., and Schiefer, G. (2006). Potential of electronic trading in complex supply chains: An experimental study. *International Journal of Production Economics*, 104(2):580–597.

Hendricks, K. B. and Singhal, V. R. (2009). Demand-supply mismatches and stock market reaction: Evidence from excess inventory announcements. *Manufacturing & Service Operations Management*, 11(3):509–524.

Hjaila, K., Puigjaner, L., Laínez, J. M., and Espuña, A. (2017). Integrated game-theory modelling for multi enterprise-wide coordination and collaboration under uncertain competitive environment. *Computers & Chemical Engineering*, 98:209–235.

Hofstra, N. and Spiliotopoulou, E. (2022). Behavior in rationing inventory across retail channels. *European Journal of Operational Research*, 299(1):208–222.

Hu, Y., Guan, Y., and Liu, T. (2011). Lead-time hedging and coordination between manufacturing and sales departments using nash and stackelberg games. *European Journal of Operational Research*, 210(2):231–240.

Hurwicz, L. (1973). The design of mechanisms for resource allocation. *The American Economic Review*, 63(2):1–30.

Kaihara, T. (2001). Supply chain management with market economics. *International Journal of Production Economics*, 73(1):5–14.

Karabati, S. and Yalçin, Z. B. (2014). An auction mechanism for pricing and capacity allocation with multiple products. *Production and Operations Management*, 23(1):81–94.

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

Karabuk, S. and Wu, S. D. (2005). Incentive schemes for semiconductor capacity allocation: A game theoretic analysis. *Production and Operations Management*, 14(2):175–188.

Katok, E. and Villa, S. (2022). Centralized or decentralized transfer prices: A behavioral approach for improving supply chain coordination. *Manufacturing & Service Operations Management*, 24(1):143–158.

Kim, B., Kim, S. W., Iravani, S. M., and Park, K. S. (2021). Simple mechanisms for sequential capacity allocations. *Production and Operations Management*, 30(9):2925–2943.

Kim, B. S. and Chung, B. D. (2017). Affinely adjustable robust model for multiperiod production planning under uncertainty. *IEEE Transactions on Engineering Management*, 64(4):505–514.

Kouvelis, P. and Lariviere, M. A. (2000). Decentralizing cross-functional decisions: Coordination through internal markets. *Management Science*, 46(8):1049–1058.

Kristensen, J. and Jonsson, P. (2018). Context-based sales and operations planning (s&op) research. *International Journal of Physical Distribution & Logistics Management*, 48(1):19–46.

Lahloua, N., Barkany, A. E., and Khalfi, A. E. (2018). Sales and operations planning (s&op) concepts and models under constraints: Literature review. *International Journal of Engineering Research in Africa*, 34(January):171–188.

Lai, G. and Xiao, W. (2018). Inventory decisions and signals of demand uncertainty to investors. *Manufacturing & Service Operations Management*, 20(1):113–129.

Lapide, L. (2004). Sales and operations planning part ii: Enabling technology. *Journal of Business Forecasting*, 23(3):18–20.

Lau, H.-S. and Lau, A. H.-L. (1996). The newsstand problem: A capacitated multiple-product single-period inventory problem. *European Journal of Operational Research*, 94(1):29–42.

Lee, H. L., Padmanabhan, V., and Whang, S. (1997). Information distortion in a supply chain: The bullwhip effect. *Management science*, 43(4):546–558.

Lee, W. J. (1994). Optimal order quantities and prices with storage space and inventory investment limitations. *Computers & Industrial Engineering*, 26(3):481–488.

Li, J., Cai, X., and Liu, Z. (2017). Allocating capacity with demand competition: Fixed factor allocation. *Decision Sciences*, 48(3):523–560.

Li, J., Min, K. J., Otake, T., and Voorhis, T. V. (2008). Inventory and investment in setup and quality operations under return on investment maximization. *European Journal of Operational Research*, 185(2):593–605.

Lin, J. T., Wu, C.-H., Chen, T.-L., and Shih, S.-H. (2011). A stochastic programming model for strategic capacity planning in thin film transistor-liquid crystal display (tft-lcd) industry. *Computers & Operations Research*, 38(7):992–1007.

Lu, L. X. and Lariviere, M. A. (2012). Capacity allocation over a long horizon: The return on turn-and-earn. *Manufacturing & Service Operations Management*, 14(1):24–41.

Luss, H. and Gupta, S. K. (1975). Allocation of effort resources among competing activities. *Operations Research*, 23(2):360–366.

Maggioni, F., Cagnolari, M., and Bertazzi, L. (2019). The value of the right distribution in stochastic programming with application to a newsvendor problem. *Computational Management Science*, 16:739–758.

Mallik, S. and Harker, P. T. (2004). Coordinating supply chains with competition: Capacity allocation in semiconductor manufacturing. *European Journal of Operational Research*, 159(2 SPEC. ISS.):330–347.

Malone, T. W. (2004). Bringing the market inside. *Harvard Business Review*, 82(4):106–14, 142.

McAdams, D. and Malone, T. W. (2011). Internal markets for supply chain capacity allocation. *SSRN Electronic Journal*.

McAfee, R. P. and McMillan, J. (1987). Auctions and bidding. *Journal of economic literature*, 25(2):699–738.

Milgrom, P. (2004). Putting auction theory to work. *Putting Auction Theory to Work*, pages 1–368.

Mishra, B. K., Raghunathan, S., and Yue, X. (2007). Information sharing in supply chains: Incentives for information distortion. *IIE Transactions*, 39(9):863–877.

Nahmias, S. and Schmidt, C. P. (1984). An efficient heuristic for the multi-item newsboy problem with a single constraint. *Naval Research Logistics Quarterly*, 31(3):463–474.

Naraghi, S. G. and Jiang, Z. (2023). Stochastic optimization of agrochemical supply chains with risk management. In Kokossis, A. C., Georgiadis, M. C., and Pistikopoulos, E., editors, *33rd European Symposium on Computer Aided Process Engineering*, volume 52 of *Computer Aided Chemical Engineering*, pages 3337–3343. Elsevier.

Narahari, Y., Narayanam, R., Garg, D., and Prakash, H. (2009). Foundations of mechanism design. In *Game Theoretic Problems in Network Economics and Mechanism Design Solutions*, pages 1–131. Springer.

Nemati, Y. and Alavidoost, M. H. (2019). A fuzzy bi-objective milp approach to integrate sales, production, distribution and procurement planning in a fmcg supply chain. *Soft Computing*, 23(13):4871–4890.

Ning, C. and You, F. (2017). A data-driven multistage adaptive robust optimization framework for planning and scheduling under uncertainty. *AIChE Journal*, 63(10):4343–4369.

Oerke, E.-C. and Dehne, H.-W. (2004). Safeguarding production—losses in major crops and the role of crop protection. *Crop Protection*, 23(4):275–285.

Oliva, R. and Watson, N. (2011). Cross-functional alignment in supply chain planning: A case study of sales and operations planning. *Journal of Operations Management*, 29(5):434–448.

Patriksson, M. (2008). A survey on the continuous nonlinear resource allocation problem. *European Journal of Operational Research*, 185(1):1–46.

Pereira, D. F., Oliveira, J. F., and Carravilla, M. A. (2020). Tactical sales and operations planning: A holistic framework and a literature review of decision-making models. *International Journal of Production Economics*, 228(July 2019):107695.

Rajan, M. V. and Reichelstein, S. (2004). A perspective on "asymmetric information, incentives and intrafirm resource allocation". *Management Science*, 50(12):1615–1623.

Rius-Sorolla, G., Maheut, J., Estellés-Miguel, S., and Garcia-Sabater, J. P. (2020). Coordination mechanisms with mathematical programming models for decentralized decision-making: a literature review. *Central European Journal of Operations Research*, 28(1):61–104.

Romano, P. (2003). Co-ordination and integration mechanisms to manage logistics processes across supply networks. *Journal of Purchasing and Supply Management*, 9(3):119–134.

Samuelson, W. (2001). Auctions in theory and practice. *Game Theory and Business Applications*, pages 295–338.

Satterthwaite, M. A. (1979). Strategy-proof allocation mechanisms. In *Game Theory*, pages 232–237. Springer.

Schlapp, J., Fleischmann, M., and Sonntag, D. (2022). Inventory timing: How to serve a stochastic season. *Production and Operations Management*, 31(7):2891–2906.

Schroeder, R. G. and Krishnan, R. (1976). Return on investment as a criterion for inventory models. *Decision Sciences*, 7(4):697–704.

See, C. T. and Sim, M. (2010). Robust approximation to multiperiod inventory management. *Operations Research*, 58(3):583–594.

Shah, N. (2004). Pharmaceutical supply chains: Key issues and strategies for optimisation. *Computers and Chemical Engineering*, 28(6-7):929–941.

Shapiro, A. (2011). A dynamic programming approach to adjustable robust optimization. *Operations Research Letters*, 39(2):83–87.

Sparks, T. C. and Lorsbach, B. A. (2017). Perspectives on the agrochemical industry and agrochemical discovery. *Pest management science*, 73(4):672–677.

Susarla, N. and Karimi, I. A. (2018). *Integrated production planning and inventory management in a multinational pharmaceutical supply chain*, volume 41, pages 551–567. Elsevier B.V., 1 edition.

Thomé, A. M. T., Scavarda, L. F., Fernandez, N. S., and Scavarda, A. J. (2012). Sales and operations planning: A research synthesis. *International Journal of Production Economics*, 138(1):1–13.

Torabi, S. A. and Hassini, E. (2009). Multi-site production planning integrating procurement and distribution plans in multi-echelon supply chains: An interactive fuzzy goal programming approach. *International Journal of Production Research*, 47(19):5475–5499.

Tuomikangas, N. and Kaipia, R. (2014). A coordination framework for sales and operations planning (s&op): Synthesis from the literature. *International Journal of Production Economics*, 154:243–262.

Ventura, J. A. and Klein, C. M. (1988). A note on multi-item inventory systems with limited capacity. *Operations Research Letters*, 7(2):71–75.

Vickrey, W. (1961). Counterspeculation, auctions, and competitive sealed tenders. *The Journal of Finance*, 16(1):8.

Vohra, R. V. (2012). Optimization and mechanism design. *Mathematical programming*, 134:283–303.

Wallace, T. F. and Stahl, R. A. (1999). *Sales and Operations Planning: The How-To Handbook*. T.F. Wallace Company.

Yanıkoğlu, I., Gorissen, B. L., and den Hertog, D. (2019). A survey of adjustable robust optimization. *European Journal of Operational Research*, 277(3):799–813.

Zhang, B. (2012). Multi-tier binary solution method for multi-product newsvendor problem with multiple constraints. *European Journal of Operational Research*, 218(2):426–434.

Zhang, B., Xu, X., and Hua, Z. (2009). A binary solution method for the multi-product newsboy problem with budget constraint. *International Journal of Production Economics*, 117(1):136–141.

Ziegler, H. (1982). Solving certain singly constrained convex optimization problems in production planning. *Operations Research Letters*, 1(6):246–252.

Curriculum Vitae

Christoph Löffel

Education

2020 - 2026	Doctoral Studies in Business Administration (Dr. rer. pol.), University of Mannheim, Mannheim, Germany
2016 - 2018	Mannheim Master in Management (M.Sc.), University of Mannheim, Mannheim, Germany
2015	Exchange Semester, San Diego State University, San Diego, USA
2013 - 2016	Business Administration (B.Sc.), University of Mannheim, Mannheim, Germany