Université de Montréal

Strategic Capacity Planning and Pricing: A Choice-based Approach

par

David Leonardo Pinzon Ulloa

Département d'informatique et de recherche opérationnelle
Faculté des arts et des sciences

Thèse présentée en vue de l'obtention du grade de
Philosophiæ Doctor (Ph.D.)
en Informatique

March 3, 2025

Université de Montréal

Faculté des arts et des sciences

Cette thèse intitulée

Strategic Capacity Planning and Pricing: A Choice-based Approach

présentée par

David Leonardo Pinzon Ulloa

a été évaluée par un jury composé des personnes suivantes :

Kim Yu
(président-rapporteur)

Emma Frejinger
(directrice de recherche)

Margarida Carvalho
(membre du jury)

Shadi Sharif Azadeh
(examinateur externe)

Benoit Perron
(représentant du doyen de la FESP)

Résumé

Cette thèse étudie les problèmes de décision stratégiques abordés par un fournisseur de services logistiques (FSL) souhaitant optimiser ses profits ou ses pertes, lorsque l'information dont il dispose à propos de la demande de ses clients pour de nouveaux services est incomplète. Nous adoptons l'hypothèse que la demande est issue de la maximisation d'utilité par les clients. Puisque la connaissance des préférences des clients par le FSL est incertaine, celles-ci sont décrites au moyen de modèles d'utilité aléatoires. La thèse est constituée de trois articles dans lesquels les problèmes traités par le FSL sont exprimés sous forme de programmes stochastiques bi-niveaux où le FSL est le leader et les clients sont les suiveurs. Les articles proposent des reformulations à un seul niveau fondées sur les propriétés duales des solutions optimales et faisant usage de la méthode d'approximation par moyenne échantillonnale pour le calcul des utilités espérées. Ces reformulations sous-tendent la construction, d'une part, de méthodes de résolution asymptotiquement exactes dont la vitesse est supérieure à celle des méthodes de pointe actuelles et, d'autre part, de méthodes heuristiques dont la vitesse et l'exactitude sont élevées.

Cette thèse est basée sur trois articles. Dans le *premier* article, le FSL offre aux expéditeurs des combinaisons de prix et de niveau de service de sorte à maximiser l'espérance des profits résultant de la fourniture des combinaisons sélectionnées par les expéditeurs, à l'inclusion des coûts associés à l'installation des lieux de service. Le programme du niveau inférieur concerne dans ce cas la sélection des combinaisons de prix et de service par les expéditeurs. Dans le *second* article, le FSL désire minimiser l'espérance du total de ses coûts de livraison et de fonctionnement en

offrant à ses clients de substituer la visite de points de cueillette et livraison à la livraison à domicile. Le programme du niveau inférieur concerne dans ce cas la sélection des points de cueillette et livraison ou de la livraison à domicile par les clients. Le *troisième* article introduit un procédé d'agrégation des scénarios dans la reformulation développée dans le premier article, permettant ainsi d'accroître la vitesse de calcul de plusieurs ordres de grandeur. En résumé, cette thèse fait avancer l'état de l'art sur les formulations stochastiques bi-niveaux pour les problèmes de localisation et de tarification. Ces problèmes sont difficiles à résoudre en raison des objectifs de maximisation du profit, des structures de coût complexes et des contraintes de capacité. D'un point de vue applicatif, la thèse fournit des perspectives managériales essentielles pour les fournisseurs logistiques.

Mots-clefs: décision stratégique, modèle de choix, programmation bi-niveau, programmation stochastique, logistique, emplacement des installations, tarification

Abstract

This thesis examines strategic decision-making problems addressed by a profit-maximizing or cost-minimizing logistic provider (LP) faced with the demand of its customers for new services, about which incomplete information is available. The demand is driven by the utility-maximizing customers' preferences and, from the LP's perspective, these preferences are surrounded by uncertainty. They are described with random utility maximizing models. The thesis comprises three articles where the LP's problems are expressed by stochastic bilevel programming formulations in which the LP is the leader and its customers are the followers. The articles propose single-level reformulations leveraging the properties of the dual optimal solutions and utilizing the method of sample average approximation to compute expected utilities. These reformulations yield asymptotically exact and faster than the state-of-the-art computation methods as well as heuristic, high-accuracy, high-speed computation methods.

The thesis is based on three articles. In the *first* article, the LP offers combinations of price and service level to the shippers and aims to maximize the expected profits of supplying the selected combinations, including the costs related to facility installation. The lower level concerns the selection of the combinations of price and service level by the shippers. In the *second* article, the LP aims to minimize its total expected delivery and operating costs by offering its customers to substitute the use of collection-and-delivery points for home-delivery service. The lower level concerns the selection of service points or home-delivery by the customers. The *third* article introduces an aggregation procedure in the reformulation of the first article,

thereby increasing the speed of computation by several orders of magnitude. To summarize, the thesis advances the literature on stochastic bilevel formulations for facility location and pricing problems. These problems are challenging to solve due to profit-maximizing objectives, complex cost structures, and capacity constraints. From an application perspective, the thesis offers managerial insights valuable to LPs.

Keywords: strategic decision, choice model, bilevel programming, stochastic programming, logistics, facility location, pricing.

Contents

Résumé v

Abstract vii

List of Tables xiii

List of Figures xv

List of Acronyms and Abbreviations xvii

Acknowledgements xix

Introduction 1

Chapter 1. Background 5

1.1. Logistics 5

1.2. Mixed Integer Linear Programming 7

1.3. Bilevel Programming 9

1.4. Stochastic Programming 10

1.5. Choice-based Optimization Problems 13

Chapter 2. First article. A Logistics Provider's Profit Maximization Facility Location Problem with Random Utility Maximizing Followers 15

2.1. Introduction 16

2.2. Related Work ... 18
2.3. Problem Description ... 22
2.4. Mathematical Programming Formulation ... 24
2.4.1. Preliminaries ... 24
2.4.2. Simulation-based Formulation ... 28
2.5. Solution Approaches ... 30
2.5.1. Continuous Relaxation of the Follower's Problem ... 31
2.5.2. Single-level Reduced Formulation and Heuristic ... 32
2.6. Illustrative Example ... 36
2.7. Computational Results ... 38
2.7.1. Experimental Setup ... 38
2.7.2. Comparison of Exact and Heuristic Solution Approaches ... 41
2.7.3. Impact of Model Parameters on Performance ... 42
2.7.3.1. Interpreting the Impact of the Demand Model Parameters ... 43
2.7.3.2. Impact of CFLP Parameters ... 46
2.7.3.3. Impact of Problem Size ... 47
2.8. Case Study ... 50
2.9. Conclusion and Perspectives ... 54
Chapter 3. Second article. A Pattern-based Reformulation of a Profit-maximizing Facility Location Problem with RUM Followers ... 57
3.1. Introduction ... 58
3.2. Pattern-based Reformulation ... 61
3.3. Computational Results ... 66

3.3.1. Experimental Setup 67
3.3.2. Comparison with BA 69
3.3.3. Comparison with H-A and H-S 69
3.3.3.1. Accuracy 69
3.3.3.2. Computing Times 72
3.4. Conclusions and Future Directions 73
Chapter 4. Third article. A Capacitated Collection-and-Delivery-Point Location Problem with Random Utility Maximizing Customers 77
4.1. Introduction 78
4.2. Literature Review 81
4.3. Problem Description 83
4.4. A Mixed Integer Non-Linear Formulation 86
4.5. Illustrative Example 89
4.6. Simulation-based Formulations 91
4.6.1. CAC Formulation 93
4.6.2. CAC with Scenario Aggregation 97
4.7. Computational Results 98
4.7.1. Experimental Setup 99
4.7.2. Impact of Uncertainty 102
4.7.3. Model and Algorithm Comparison 104
4.8. Case Study 107
4.9. Conclusions 112
Conclusion 115

References 119

Appendix A. First Article 131

A.1. Notation 131

A.2. Single-level MILP 132

A.3. Parameters in Experimental Results 135

A.4. Proposition in Section 2.5.2 137

A.5. Detailed Results 138

Appendix B. Second Article 143

B.1. Notation 143

B.2. Dominated Single-level Reformulation 146

B.3. Additional Results 147

Appendix C. Third Article 151

C.1. Additional Results 151

List of Tables

2.1 Summary of the most closely related work 21

2.2 Computing times for hard ($\alpha = -0.1, \beta = 1, R = 1$) and easy ($\alpha = -0.1, \beta = 0.5, R \in \{5,10\}$) classes of instances 49

2.3 Computing times for *ORLib* class instances adapted from problems *IV* to *XIII* in Beasley (1988) 51

2.4 Computing times and gaps for large instances in *ORLib* class, adapted from problem sets A and A-reduced 52

2.5 Price and service level data in the case study 53

3.1 Example of potential agreements 62

3.2 Comparison of average *GAP* obtained with BA-R and BA 71

3.3 Minimum and Maximum Absolute *GAP* between BA-R and HA 72

3.4 Average computing time BA-R, small instances with $|S| = 20{,}000$ 74

3.5 Comparison of computing times BA-R vs H-A, H-instances 75

4.1 Summary of related work 83

4.2 Choice probabilities for each CDP location decision 89

4.3 Results for location decisions x_d 90

4.4 Average computing time for S-class instances. 107

4.5 Average computing time for A-class instances 108

4.6 Average computing time for Z-class instances 109

4.7 Average computing time for D-class instances 110

4.8 Customer categories and utility function parameters . . . 110
4.9 CDP instance data . . . 112

A.1 Indexes, sets and parameters . . . 132
A.2 Stochastic data, variables and functions for model formulation . . . 133
A.3 Possible values of the configuration parameter α . . . 136
A.4 Values for the scale parameter β . . . 136
A.5 Price level sets for the *ORlib* class of instances . . . 137
A.6 Relative differences between $Z_{10\%}$ and $Z_{90\%}$ for optimal solutions with *BA* 139
A.7 Average computing times for *BA* and *H-S* and *H-A* for each instance configuration . . . 140
A.8 Average optimality gaps between *BA* and *H-S* and *H-A* . . . 141

B.1 Indexes, sets and deterministic parameters . . . 144
B.2 Derived and decision variables . . . 145
B.3 Computing time CAC vs Standard . . . 148

C.1 Relative differences between $Z_{10\%}$ and $Z_{90\%}$ for optimal solutions with BA-R . . . 152

List of Figures

2.1 Overview of the illustrative example setup . . . 37
2.2 Results on relative gap and entropy . . . 43
2.3 Optimal expected profit and computing time according to α . . . 44
2.4 Optimal expected profit and computing time according to β . . . 45
2.5 Optimal profit according to α for $\beta = 1$, $\beta = 32$ and a deterministic utility 45
2.6 Optimal expected profit and computing time according to ratio between total capacity and total demand R . . . 47
2.7 Demand zones and terminals in the case study . . . 52
3.1 Ratio of computing times with BA-R and BA, given $|I|, |J|$ and entropy in [a,b) . . . 70
3.2 Ratio of computing times with BA-R and BA, given $|I|, |J|$ and $|S|$ . . . 70
4.1 Zones and CDPs . . . 89
4.2 Number of open CDPs in the optimal solution for different values of the ratio between entropy and max entropy for each instance . . . 103
4.3 Distribution of the optimal number of CDPs . . . 104
4.4 Distribution of the optimal expected cost . . . 105
4.5 Postal Codes and CDPs. . . . 111

B.1 Distribution of optimal fraction captured by CDPs . . . 147
B.2 Distribution of the relative performance in low entropy instances . . . 149

List of Acronyms and Abbreviations

AP	Automated Partition
CAC	Closest Assignment Constraints
CBFLP	Choice-Based Facility Location Problem
CFLP	Capacitated Facility Location Problem
IP	Integer Programming
LP	Logistics Provider
LRP	Location and Routing Problem
MILP	Mixed Integer and Linear Programming
RUM	Random Utility Maximizing
SAA	Sample Average Approximation
UP	User Partition

Acknowledgements

To my beloved wife, Solange, for her long-lasting support, patience, and love. To my adored kids, Maite, Gabriel, and Nicolas, my stress relievers and a great source of joy and happiness. To my parents, Livio and Josefina, for their unconditional affection and motivation. To my brothers Eduardo and Raul and my sister Romina, I miss you all.

I also want to thank my supervisor, Emma Frejinger, for her outstanding academic support, mentorship, and continuous encouragement. To Bernard Gendron, for trusting me while we worked together. I am deeply grateful to Eric Larsen and Ammar Metnani, who helped me improve this work.

I gratefully acknowledge the financial support I received from the Ecuadorian Secretaría de Educación Superior, Ciencia, Tecnología e Innovacion (SENESCYT) under its international scholarships program and from University Escuela Superior Politécnica del Litoral, Ecuador.

Introduction

Logistics providers (LP), the backbone of modern supply chains, are crucial in ensuring that goods move efficiently from suppliers to consumers. They play a significant role in the Canadian economy, with an expected contribution of around CAD 106.4 billion in 2024 to the GDP (Intelligence, 2024). The LPs – or third-party logistics (3PL) – are essential for moving goods across vast areas. However, they must make decisions in increasingly challenging environments, characterized by disruptions due to global events, technological integration, increased and complex customer expectations, difficult capacity management (Chopra and Meindl, 2016). Natural disasters and pandemics, such as COVID-19, expose supply chains' vulnerabilities, leading to increasing costs and delivery delays. Integrating new technologies, such as artificial intelligence and blockchain, is essential for optimizing performance but also requires significant investments. Customers with increased expectations demand that the LPs provide higher service quality levels, such as flexible and faster deliveries, which implies higher operational costs. However, not all these customers are willing to pay higher price levels. These stringent requirements may lead to failed 3PL partnerships (see, e.g., Logistics, 2022; Melvyn Peters and Randall, 1998). In this context, managing capacity is challenging as it requires strategic decisions involving significant investments that, given the fluctuating and uncertain demand, can be overestimated, resulting in a loss of efficiency and profitability. The combination of some or all these factors contributes to creating a complex landscape and requires LPs to make high-quality strategic decisions to remain competitive.

The LPs seek attractive solutions that improve their financial results and allow them to adapt to constantly changing and more complex demand behaviors, especially those arising from individual choices. To this end, the LPs must determine (i) what they offer, including, inter alia, the service and prices for their customers and (ii) the logistical capacities they need to fulfill the customers' requirements. It is critical that the demand is sensitive to the LP's offers. Therefore, the LPs must understand that the interplay between demand and supply is pivotal in accomplishing their goals and set their decision-making approach accordingly.

Choice models are the primary methods for representing the demand behavior at disaggregated levels (Ricard and Bierlaire, 2024), whereas mathematical programming is used to model supply decisions. Several researchers have integrated choice models in optimization frameworks (see e.g., Haase and Müller, 2014; Pacheco Paneque et al., 2021; Mai and Lodi, 2020). This has been accomplished in applications to facility location, transportation, and revenue management problems.

In this thesis, we study choice-based optimization problems for strategic decisions involving capacitated facility locations, service levels, and pricing. Moreover, we focus on problems where the choice (or preferences) of individuals or groups of individuals with similar characteristics is not well-known and, hence, uncertain. For this purpose, we study bilevel formulations that include random utility maximizing followers.

The following outlines this thesis' structure and specific contributions.

Chapter 1 introduces the main mathematical concepts and results underlying the three articles comprised in the thesis. The *first* article, in Chapter 2, addresses a problem of high practical relevance to LPs in general and to our industrial partner, Purolator, in particular. This article proposes a high-quality heuristic addressing a stochastic bilevel facility location and pricing problem. The heuristic approach is based on a reduced, single-level, deterministic reformulation exploiting the properties of dual optimal solutions and utilizing the sample average approximation for the calculation of expected utilities. The problem is challenging due to the complexities resulting from the interplay between the LP, the shippers, and the shippers' customers. We report an extensive experimental study where we analyze the impact of

model parameters on solutions and computing times. The *second* article, in Chapter 4, addresses a capacitated collection-and-delivery-point location problem where there is uncertainty about customers' locations and preferences. The problem presents two challenging aspects: (i) potentially conflicting objectives between the LP and its customers, and (ii) strict assignments of customers to facilities. We introduce formulations that can be efficiently solved using a standard Benders decomposition method implemented with general-purpose solvers. These formulations are based on closest assignment constraints (CAC) and scenario aggregation. We show that the CAC reformulation corresponds to a single-level transformation of a bilevel program where we use an alternative formulation for the follower's subproblem. We present an extensive computational study analyzing the computing times and the effects of uncertainty on the solutions. We also demonstrate the practical applicability of our work through a realistic case study about our industrial partner. The *third* article, in Chapter 3, proposes an asymptotically exact method based on the original simulation-based formulation introduced in the first article. Through the aggregation of scenarios featuring similar agreement patterns, this reformulation leads to sharp improvements in computing time. Finally, the general conclusion summarizes the thesis and discusses limitations and directions for future research.

To summarize, the thesis advances the literature on stochastic bilevel formulations for facility location and pricing problems. These problems are challenging to solve due to profit-maximizing objectives, complex cost structures, and capacity constraints. From an application perspective, the thesis offers managerial insights valuable to LPs.

Chapter 1

Background

This chapter provides a background of logistics and outlines the core mathematical concepts and results used in the thesis. Section 1.1 briefly describes the logistic market used as a reference in this work. Section 1.2 introduces continuous and mixed integer linear programming. The latter are instrumental to the definition and transformation of bilevel programs, overviewed in Section 1.3, into single-level formulations. Section 1.4 presents the mathematical programming models that account for the stochasticity resulting in our context from incomplete information about preferences. These play a pivotal role in our work. Finally, Section 1.5 outlines the class of choice-based optimization problems where our work is situated.

1.1. Logistics

Logistics involves processes related to planning, executing, and controlling the movement of goods, services, and information from manufacturers to consumers (or customers). It includes activities that streamline storage, transportation, inventory management, and information flow, ensuring the right products reach customers while respecting delivery times. In essence, it aims to deliver products effectively while minimizing costs. The logistics market is valued at USD 3,794 billion in 2023 (Grand View Research, 2023). The sector encompasses three main stakeholders: LPs, shippers, and customers. The dynamics of the market, driven by economic trends,

technological innovations, and regulatory obligations, impact the interactions among these stakeholders.

This thesis focuses on specific problems faced by LPs. The LPs are responsible for facilitating the movement, storage, and distribution of goods. They have evolved over time increasing the range of services provided and the engagement in supply chain management. They exist to help other companies (their clients) streamline key processes of the supply chain. They aim to improve the quality of service and profitability of their clients, while staying profitable themselves. This is challenging because LPs operate within a highly competitive market where factors like cost-efficiency, service quality, and technological innovation influence their success. As examples of leading LPs, we can cite Fedex, DHL, UPS, Purolator.

Shippers are exporters, manufacturers, wholesalers, or retailers that need to transport goods from their facilities to distribution centers or end customers. Many shippers plan and operate their own logistics resources, such as delivery trucks and warehouses. However, a great share of these shippers (see Placek, 2022) opts for outsourcing transportation and warehousing to reduce costs, improve customer service, and increase efficiency. This allows shippers to concentrate on their core business while more specialized firms (LPs) engage in outsourced activities (Abbasi et al., 2024). In this context, shippers choose to outsource when they anticipate cost savings and/or enhanced service quality. Otherwise, they may forgo partnerships with LPs. Service quality primarily refers to the experience and satisfaction of end customers.

In the logistics sector, customers comprise both businesses and individual consumers. They typically expect delivery services to be fast, cost-effective, and reliable. The surge in e-commerce has dramatically influenced these expectations, with companies such as Amazon raising the bar for same-day or next-day delivery. Consequently, LPs and shippers have been compelled to enhance their operations to fulfill these complex and increasing requirements, sometimes by agreeing to work together.

1.2. Mixed Integer Linear Programming

A linear program is an optimization problem where a decision vector $x \in \mathbb{R}^n$ must minimize a linear objective function $c^\intercal x$ while satisfying a set of linear constraints $Ax \geq b$, where $A \in \mathbb{R}^{m \times n}$ and $c \in \mathbb{R}^n$. In standard form, a linear program is stated as:

$$\begin{aligned} &\min_{\boldsymbol{x}} \ c^\intercal x \\ &\text{s.t } Ax = b \\ &\quad x \geq 0. \end{aligned} \tag{1.1}$$

Program (1.1) can be viewed as the primal representation of a the continuous linear program. The latter is associated with a corresponding dual representation, given by:

$$\begin{aligned} &\max_{\boldsymbol{y}} \ b^\intercal y \\ &\text{s.t } A^\intercal y \leq c \\ &\quad y \in \mathbb{R}^m. \end{aligned} \tag{1.2}$$

A fundamental property of linear programming lies in the close relation between Programs (1.1) and (1.2) (Vanderbei, 2020). In particular, strong duality states that if an optimal solution exists for Program (1.1), then the optimal objective values of both programs are the same. More precisely, let x^* and y^* be the corresponding optimal solutions for Programs (1.1) and (1.2), then $c^\intercal x^* = b^\intercal y^*$. Therefore, Program (1.1) can be written as the following set of constraints:

$$
\begin{aligned}
c^\intercal x &= b^\intercal y \\
Ax &= b \\
A^\intercal y &\leq c \\
x \geq 0, y &\in \mathbb{R}^m.
\end{aligned} \tag{1.3}
$$

Linear programming can be extended to handle mixed integer linear programming (MILP) problems. In this case, $x \in \mathbb{R}^{n-p} \times \mathbb{Z}^p$. A particular class of such problems is that where $p = n$. These are called integer linear programming (ILP) problems. Linear programming is a special case of a broader class of optimization, that of convex optimization. Problems of this class feature a convex objective function and their constraints define a convex set. More precisely, $\min f(x) : x \in X$ is a convex optimization problem if $\forall x_1, x_2 \in X$, $\theta \in [0,1]$ and $\alpha \in [0,1]$:

$$f(\alpha x_1 + (1-\alpha)x_2) \leq \alpha f(x_1) + (1-\alpha) f(x_2)$$

$$\theta x_1 + (1-\theta)x_2 \in X.$$

Note that the objective function f must not necessarily be linear to be convex. Also, the set X does not require that the associated constraints are all linear to be convex. However, linearity is a sufficient condition for convexity.

Several pivotal optimization problems in operations research lie in the MILP class and are therefore NP-hard. MILP problems are not convex as their sets of feasible solutions are not convex. However, a desirable property of a number of MILP problems, total unimodularity (TU) of the matrix A, allows the relaxation of the integrality constraints (Wolsey, 1998). These particular MILP problems can be treated as convex, continuous linear programming problems, thus making them easier to solve using the well-known *simplex* or *interior-point* methods. Moreover, when A satisfies TU, a MILP problem can benefit from the properties of strong duality, which do not hold for general MILP problems. When the TU property does not hold for $P = \{x \in \mathbb{Z}^n : Ax \leq b\}$, it is customary to start from the linear relaxation $H = \{x \in \mathbb{R}^n : Ax \leq b\}$, of the integer problem and to progressively

explore restrictions of this domain until an optimal solution to the original problem is found.

1.3. Bilevel Programming

A particular case, occurring in the problems addressed in this thesis, arises when the set of feasible solutions includes an embedded optimization problem. Bilevel programs are a class of non-convex hierarchical optimization problems with two decision-makers, a leader and a follower, who seek to optimize their objective functions. For further details, readers are referred to the textbooks Bard (1998), Dempe (2002), and Dempe and Zemkoho (2020). The decision-making process is not simultaneous, as the follower reacts to the decision made by the leader in the first place. Bilevel programs are generally described as follows:

$$\min_{x \in X, y \in Y} \; F(x,y) \tag{1.4a}$$

$$\text{s.t. } G(x,y) \geq 0 \tag{1.4b}$$

$$y \in \arg\min_{y \in Y} \{ f(x,y) : g(x,y) \geq 0 \}. \tag{1.4c}$$

In the bilevel Program (1.4), (1.4c) constitutes the follower's problem. Decision variables x and y are controlled by the leader and follower, respectively. The objective function $F(x,y)$ and the constraints $G(x,y) \geq 0$ relate to the leader, whereas $f(x,y)$ and $g(x,y) \geq 0$ relate to the follower.

It frequently occurs that the optimal solution set of the follower's problem is not a singleton. We distinguish two ensuing approaches in these circumstances: optimistic and pessimistic. These respectively hinge on the postulates of presence and absence of cooperation between the leader and the follower. In the optimistic approach, the follower is assumed to cooperate by choosing among its optimal solutions the one that maximizes the leader's objective function. This approach is more tractable than the pessimistic one, and most algorithms for solving bilevel problems adopt it. Throughout this thesis, we also apply the optimistic approach.

Bilevel problems are challenging to solve: even in the linear case, they belong to the NP-hard class (Hansen et al., 1992). Yet, there have been advances in the methods proposed for their solution (Dempe and Zemkoho, 2020). This thesis follows the single-level reformulation approach which is applicable when the follower's problem is a continuous or integer linear program satisfying the TU property. This approach relies on the reformulation of the follower's problem using the strong duality theorem. Thus, for a mixed integer bilevel program given by (1.5), the corresponding single-level reformulation using strong duality theory is expressed by (1.6). In Program (1.6) π identifies the dual variables associated with constraints $Cx + Dy \geq e$. Also, Constraints (1.6c), (1.6e), (1.6d) correspond to the strong duality conditions applied to the follower's Problem (1.5c).

$$\min_{x} \; c^\intercal x + f^\intercal y \tag{1.5a}$$
$$\text{s.t. } Ax + By \geq d \tag{1.5b}$$
$$y \in \arg\min_{y}\{g^\intercal y : Cx + Dy \geq e\}. \tag{1.5c}$$

$$\min_{x,y,\pi} \; c^\intercal x + f^\intercal y \tag{1.6a}$$
$$\text{s.t. } Ax + By \geq d \tag{1.6b}$$
$$g^\intercal y = (e - Cx)^\intercal \pi \tag{1.6c}$$
$$Cx + Dy \geq e \tag{1.6d}$$
$$D^\intercal \pi \leq g. \tag{1.6e}$$

1.4. Stochastic Programming

The methods overviewed in the previous sections belong to the realm of deterministic optimization. They postulate that perfect information is available to the decision-makers. However, in real-world decision-making, limited or incomplete information leads to uncertainty about future events or conditions. This uncertainty significantly limits the effectiveness of deterministic optimization approaches as it

spreads through the entire decision-making process. For instance, uncertainty is encountered in relation to demand, prices, and availability of resources.

To deal with uncertainty, we can structure optimization problems as single, two-stage, and multi-stage model formulations (Birge and Louveaux, 2011). In single-stage formulations, a decision is implemented without considering any recourse. In contrast, in multi-stage formulations, decisions are made progressively: decisions that occurred in previous stages impact the choices available in the current stage. The two-stage approach lies between those polar cases. In this case, we distinguish the first-stage decisions made before uncertainty has been resolved from the second-stage decisions made after the uncertainty is fully resolved. This two-stage approach is suitable for strategic problems where initial decisions are less prone to adjustments and set the stage for future decisions that depend on uncertain outcomes.

The literature proposes several approaches for handling two-stage stochastic programs. These approaches include robust optimization, distributionally robust optimization, and stochastic programming. Interested readers are referred to the textbook Birge and Louveaux (2011) for a detailed presentation. Robust optimization addresses the worst cases in a set of uncertain outcomes. It aims to find the solutions protecting against the most adverse scenario(s). Distributionally robust optimization can be viewed as a refinement of the previous approach. In this case, a family of probability distributions describes the uncertainty and the aim is to find solutions that perform well under the worst-case distribution of uncertain parameters.

Stochastic programming assumes that a probability distribution describing the uncertainty surrounding certain parameters is known or can be estimated from historical data. It aims to find the optimal solution to a program whose objective is usually the expectation of a gain or a loss. In our case, it is reasonable to assume that estimates of the relevant distributions can be calculated from historical data, whence stochastic programming can be viewed as a tractable approach.

In the simplest linear case, two-stage stochastic programming problems can be expressed as (1.7) where x and $y(w)$ respectively denote the first-stage and second-stage decision variables. Also, T, W, and h are respectively known as the recourse

matrix, the technological or coupling matrix, and the resource vector:

$$\min\ z = c^{\intercal}x + \mathcal{D}(x) \tag{1.7a}$$

$$\text{s.t. } Ax = b, \tag{1.7b}$$

$$x \geq 0. \tag{1.7c}$$

where

$$\mathcal{D}(x) = \mathbb{E}_{\xi}[\mathcal{Q}(x,\xi(w))] \tag{1.7d}$$

$$\mathcal{Q}(x,\xi(w)) = \min\{q(w)^{\intercal}y(w) : T(w)x + Wy(w) = h(w), y(w) \geq 0\}. \tag{1.7e}$$

In equation (1.7e), $\mathcal{Q}(x,\xi(w))$ is the second stage value function, given x and a realization w of the random vector $\xi(w) = (q(w), h(w), T(w))$.

In general, the main difficulty with two-stage stochastic programming lies in the non-linearity associated with $\mathcal{D}$, the recourse function. Monte Carlo simulation methods constitute alternative approaches to address these stochastic programs (Birge and Louveaux, 2011). The most widely applied and documented among the latter is the external sampling approach known as sample average approximation (SAA) (see Shapiro, 2003, for more details). Its main idea is to replace the recourse function $\mathcal{D}(x)$ with a Monte Carlo estimate

$$\mathcal{D}^{|S|}(x) = \frac{1}{|S|}\sum_{s\in S}\mathcal{Q}(x,\xi_s), \tag{1.8}$$

where S is a finite sample set of observations s for the random vector ξ. Therefore, Program (1.7) is approximated by a deterministic formulation given by Program (1.9):

$$\min\ \hat{z} = c^{\intercal}x + \frac{1}{|S|}\sum_{s\in S} q_s^{\intercal}y_s \tag{1.9a}$$

$$\text{s.t. } Ax = b, \tag{1.9b}$$

$$T_s x + Wy_s = h_s, s \in S, \tag{1.9c}$$

$$x \geq 0, y_s \geq 0, s \in S. \tag{1.9d}$$

By using Program (1.9), it is possible to take advantage of existing optimization techniques for MILP. One of the well-known approaches in this regard is an adaptation of the Benders decomposition algorithm to stochastic programs, the L-shaped method (Van Slyke and Wets, 1969; Birge and Louveaux, 2011). The L-shaped method leverages the structure of Program (1.9) by decomposing it into a master problem and $|S|$ subproblems. This method assumes that the random vector ξ has (or can be well-approximated by a) finite support and constructs a piecewise linear approximation to the recourse function. The master problem is updated in each iteration of the algorithm. Note that although L-shaped is an exact method, the solution resulting from the introduction from the SAA and possibly also the approximation through a finite support is approximate. However, as an application of the law of large numbers, this approximate solution converges stochastically to the true optimal value as $|S|$ increases. Bazotte et al. (2024) focus on SAA for models with endogenous uncertainty.

1.5. Choice-based Optimization Problems

Choice-based optimization problems address the necessity of integrating the interactions between demand at the disaggregated level and the related decision-making process on the supply side (Pacheco Paneque et al., 2021; Ricard and Bierlaire, 2024). This framework embeds choice models, particularly those known as random utility maximization (RUM) models, into optimization problems. Due to its simplicity and convenience, and despite its limitations, the multinomial logit (MNL) model is among the latter the most frequently used in the literature on choice-based optimization. Other well-known choice models are the mixed MNL (also known as logit mixture) and the nested logit. For details on discrete choice models, readers can refer to Train (2009); Ben-Akiva and Lerman (1985).

One of the main challenges of incorporating choice models is the resulting complex optimization model. However, some strategies have been devised to address this issue. Linearization of the logit model (Haase and Müller, 2014) and simulation of linear utilities relying on SAA (Pacheco Paneque et al., 2021) are the most common.

Most of the applications of choice-based optimization problems are related to facility location and urban transportation problems (Ricard and Bierlaire, 2024). A vast literature in choice-based facility location problems has been developed (e.g., Mai and Lodi, 2020; Ljubić and Moreno, 2018; Legault and Frejinger, 2023). This problem is, in general, formulated as

$$\max_{x \in X} z = \sum_{i \in I} \sum_{j \in J} d_j P(i,j|i \in I_x), \tag{1.10}$$

where $P(i,j|i \in I_x)$ denotes the probability that customers located in j are captured by the opened facility $i \in I_x$, and d_j the demand of the corresponding customers. In (1.10), I_x denotes the set of opened facilities, and $x \in X$, a set of feasible decisions for the location of the facilities. To illustrate the intrinsic nonlinearity in (1.10), let us assume a logit model holds. Then, $P(i,j|i \in I_x)$ is computed with

$$\max_{x \in X} z = \sum_{i \in I} \sum_{j \in J} d_j \sum_{i \in I} \frac{e^{v_{ij}} x_i}{\sum_{i' \in I} e^{v_{i'j}} x_{i'}}, \tag{1.11}$$

where v_{ij} denotes the deterministic part of utility of location i for customers located in j, and x_i identifies the decision of open a facility in location i ($x_i = 1$) or not ($x_i = 0$).

In this thesis, we examine choice-based optimization problems involving facility location and pricing decisions. To solve these optimization problems, we approximate the choice probabilities by simulating utilities, as proposed in Pacheco Paneque et al. (2021). To account for the conflicting objectives of the decision-makers, we formulate the overall optimization problem as a bilevel program.

Chapter 2 First article

A Logistics Provider's Profit Maximization Facility Location Problem with Random Utility Maximizing Followers

by

David Pinzon Ulloa[1], Emma Frejinger[1], and Bernard Gendron[1]

([1]) Département d'informatique et de recherche opérationnelle, Université de Montréal, Montréal, Canada.
CIRRELT, Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation

This article has been published in Computers and Operations Research.

The main contributions of the authors for this article are as follows.

- David Pinzon proposed the model formulations, designed, implemented, ran the experiments, and wrote the article.
- Bernard Gendron suggested the topic and to frame it as a choice-based optimization problem.
- Emma Frejinger provided advice on the methodology and contributed to the mathematical formulations and writing of the article.

ABSTRACT. We introduce a strategic decision-making problem faced by logistics providers (LPs) seeking facility location decisions that lead to profitable operations. The profitability depends on the revenue generated through agreements with shippers, and the costs arising when satisfying these agreements. The latter depends, in turn, on the service levels and characteristics of the shippers' customers. However, at a strategic level, the LP has imperfect information thereof.

We propose a stochastic bilevel formulation where a given LP (leader) anticipates the decisions of shippers (followers) arising from a random utility maximization model. Using a sample average approximation and properties of the associated optimal solutions, we propose a heuristic that can compute high-quality solutions. It is based on a reduced single-level mixed integer linear programming formulation that can be solved by a general-purpose solver after a preprocessing phase. Experimental results show that expected profit highly depends on shippers' price sensitivities. Underestimating price sensitivities can lead to an overestimation of the expected profit.

Keywords: Facility location, logistics, pricing, bilevel programming, profit maximization, random utility maximization models

2.1. Introduction

Growth in online shopping and omni-channel distribution leads to logistical challenges, and outsourcing of logistics operations is a widely employed strategy (Langley and Infosys, 2019). We focus on transportation services offered by a *logistics provider* (LP), i.e., a carrier or third-party logistics provider, to its clients, shippers. The market environment for LPs is challenging. It is highly competitive and demand patterns change over time. There is also a need to reduce the negative environmental footprint

which requires changes in infrastructure and fleet composition (e.g., by electrification). LPs, therefore, face complex planning problems where it is key to consider profitability at all stages.

This work addresses the situation faced by LPs seeking to make strategic facility location decisions that lead to profitable operations. The profitability depends on the revenue generated through agreements with shippers – demand for offered service levels and associated prices – and the cost incurred when satisfying those agreements. The latter depends, in turn, on service levels and customer characteristics, such as location. Focusing on a given LP, we introduce a profit-maximizing facility location problem where the decisions, in addition to locations, pertain to service levels and prices. It is challenging to solve because of the interdependence between different actors. Namely, while the LP makes agreements with shippers, the cost of operations largely depends on the characteristics of the shippers' customers. At a strategic planning level, the LP has imperfect information thereof.

Our strategic profit maximization facility location problem has similarities with facility location and pricing problems. Closest to our work are studies dealing with both facility location and pricing decisions formulated as bilevel programs (e.g., Dan et al., 2020). These are hierarchical decision-making problems where a leader first makes decisions, anticipating the follower's reactions (lower-level problem). In our problem, a given LP is the leader, and shippers are the followers. A distinctive and challenging aspect of our problem is the endogenous uncertainty associated with the LP's cost (leader objective function) arising from unknown characteristics of both shippers and their customers. Unlike existing work (e.g., Dan et al., 2020), we consider a more complex cost structure that depends on random-utility-maximizing shippers and customers. Our work is complementary to studies on tactical transportation service procurement problems (e.g., Lafkihi et al., 2019), as the strategic facility location decisions we focus on are made before any procurement process takes place.

This paper offers four main contributions. *First*, we address a problem of high practical relevance to LPs in general, and to our industrial partner, Purolator, in particular. This problem, to the best of our knowledge, has not been studied before.

Second, we show that the problem can be naturally modeled with a stochastic bilevel formulation where shippers make decisions according to a *random utility maximization (RUM) model* (McFadden, 1981). Noteworthy is the intrinsic and endogenous uncertainty in the LP's costs due to imperfect knowledge of the shippers' reactions and their customers' characteristics. *Third*, leveraging a simulation-based formulation and properties of the associated solutions, we propose a heuristic that solves the reduced single-level mixed-integer linear formulation after a preprocessing phase. *Fourth*, we report an extensive experimental study where we analyze the impact of model parameters on solutions and computing times. For most settings, we solve instances of sizes relevant in practice within a reasonable computing time. Optimal expected profit displays important variations depending on price sensitivity and the level of uncertainty associated with shipper and customer reactions. Notably, situations where there is no profitable solution are identified quickly (i.e., expected costs exceed expected revenue). The occurrence of the latter depends on the level of price sensitivity. Moreover, expected optimal profit increases with increasing level of uncertainty. Underestimating price sensitivities can hence lead to overestimating the optimal expected profit.

The remainder of the paper is structured as follows. Section 2.2 reviews related work and Section 2.3 offers a detailed description of the problem. Section 2.4 exposes the mathematical programming formulation. Section 2.5 describes the algorithmic approach proposed, which comprises a single-level mixed integer linear program (MILP) that can be solved effectively with a general-purpose solver. Section 2.6 supplies an illustrative example, Section 2.7 reports detailed experimental results, and Section 2.8 describes a case study of our industrial partner. Finally, Section 2.9 concludes.

2.2. Related Work

There is a wealth of literature on bilevel problems with a deterministic (e.g., Colson et al., 2007) or stochastic followers (e.g., Xu et al., 2012; Sinha et al., 2016; Beck et al., 2023) problem and a comprehensive review would exceed the scope of

this paper. Hence, this section discusses related work on facility location *and* pricing, and transportation service procurement in the objective of identifying similarities and contrasts with our problem.

While extensive literature studies facility location and pricing problems separately, that treating the joint problem is relatively scarce. We may therefore distinguish facility location problems in which customers show preferences for facilities without taking pricing into account (e.g., Casas-Ramirez et al., 2018; Lee and Lee, 2012; Gutjahr and Dzubur, 2016; Mai and Lodi, 2020; Ljubić and Moreno, 2018; Küçükaydin et al., 2011), from those that include pricing in the problem (e.g., Diakova and Kochetov, 2012; Alizadeh et al., 2013; Lüer-Villagra and Marianov, 2013; Dan et al., 2020). Only a few of the latter (e.g., Dan et al., 2020) consider followers showing preferences towards price levels in addition to the characteristics of the facilities. A number of existing studies include objectives of cost minimization (see, e.g., Casas-Ramirez et al., 2018; Gutjahr and Dzubur, 2016), revenue maximization (see, e.g., Alizadeh et al., 2013; Brotcorne et al., 2000) or captured-demand maximization (see, e.g., Mai and Lodi, 2020; Ljubić and Moreno, 2018). Numerous works consider a profit-maximizing objective (e.g., Suárez-Vega et al., 2004; Küçükaydin et al., 2011; Saidani et al., 2012; Lüer-Villagra and Marianov, 2013; Panin et al., 2014; Kress and Pesch, 2016; Dan et al., 2020). However, to the best of our knowledge, the literature on bilevel pricing *and* facility location with a profit-maximizing leader is limited (e.g., Lüer-Villagra and Marianov, 2013; Panin et al., 2014; Kress and Pesch, 2016; Dan et al., 2020), and scarce if random-utility-maximizing followers are also considered (Lüer-Villagra and Marianov, 2013; Kress and Pesch, 2016).

We address a facility location and pricing problem where the objective is to maximize profit. Unlike existing work, the cost in the leader's problem depends on follower reactions that are imperfectly known. In other words, the endogenous uncertainty surrounding costs and depending on leader decisions constitutes a key challenge. In existing work (e.g., Lüer-Villagra and Marianov, 2013; Panin et al., 2014; Kress and Pesch, 2016; Dan et al., 2020), the cost is easy to incorporate, whereas in our problem, the relationship between price levels and costs is more complex. The former is associated with service levels and shippers, and the latter are associated

with the location of shippers' customers that, in turn, only impact such costs if they are included in the agreements. Our problem also differs because customers do not choose a facility. Instead, they are captive of the shippers' and LP's decisions, and their preferences, even when they are taken explicitly into account, are not related to facilities. Heuristics are commonly used to solve bilevel facility location and pricing problems with customer preferences (Lin and Tian, 2022). However, because of the distinguishing features of our problem, existing heuristics used in related work do not directly apply. For example, the genetic algorithm of Lüer-Villagra and Marianov (2013) and the greedy heuristic of Kress and Pesch (2016).

Our strategic decision-making problem presents similarities with transportation service procurement problems arising at the tactical level. This literature examines a number of mechanisms defining short- or long-term agreements between shippers and carriers (in our case LPs). Noteworthy are combinatorial auctions for transportation procurement (e.g., Lafkihi et al., 2019). Typically the focus lies on single-level formulations and first-price auctions assumptions (e.g., Lyu et al., 2021; Hammami et al., 2022). However, there are also bilevel formulations where the leader is a carrier and shippers are the followers (e.g., Yan et al., 2018). We are not aware of any work that considers facility location decisions. Our work is complementary to the literature on procurement as we focus on strategic facility location decisions and only take pricing into account at a high level. Table 2.1 summarizes the key aspects of our work and compares it with the related works.

Related Work	RUM Followers	Pricing Decisions	Service Level Decisions	Location Decisions	Bilevel Formulation	Max Profit	Endogenous Uncertainty in Costs	Carrier Shipper Agreements
Casas-Ramirez et al. (2018)				✓	✓			
Lee and Lee (2012)				✓				
Gutjahr and Dzubur (2016)				✓	✓			
Mai and Lodi (2020)	✓			✓				
Ljubić and Moreno (2018)	✓			✓				
Küçükaydin et al. (2011)	✓			✓	✓	✓		
Diakova and Kochetov (2012)		✓		✓				
Alizadeh et al. (2013)		✓			✓			
Lüer-Villagra and Marianov (2013)	✓	✓		✓		✓		
Lin and Tian (2022)		✓		✓	✓			
Kress and Pesch (2016)	✓	✓		✓	✓	✓		
Dan et al. (2020)		✓	✓	✓	✓	✓	✓	
Brotcorne et al. (2000)		✓			✓			
Panin et al. (2014)		✓		✓	✓	✓		
Lafkihi et al. (2019)		✓						✓
Lyu et al. (2021)		✓				✓		✓
Hammami et al. (2022)		✓				✓		✓
Yan et al. (2018)					✓			✓
Our work	✓	✓	✓	✓	✓	✓	✓	✓

Table 2.1. Summary of the most closely related work

2.3. Problem Description

This section provides a detailed description of the problem and introduces the relevant notation. Tables A.1 and A.2 in Appendix A.1 summarize the latter. We start by describing the actors – LP, shippers and customers – and related concepts. We then outline the decision-making problem of the LP and its link to shippers and customers.

We consider an LP that offers logistics operations services to its clients, a set of shippers N. In turn, each shipper $n \in N$ needs to satisfy the demand d_j of its customers $j \in J_n^{'} \subseteq J$, where J denotes the set of all customers, regardless the shipper. Note that we distinguish the LP's *clients* (i.e., the shippers) from the *customers* of the shippers. The customers with the same profile and preferences are grouped into categories $k \in K_n$ according to known characteristics. We denote by $J_k \subseteq J$ the set of customers who belong to the same category $k \in K_n$. They can be end-consumers or businesses depending on the type of business the LP deals with – business-to-consumer (B2C) or business-to-business (B2B). Note that in this case, $J_n^{'} \cap J_k$ identifies the subset of customers of shipper n and customers in category k; therefore, if $k \notin K_n$, $J_n^{'} \cap J_k = \emptyset$.

The LP seeks to make strategic facility location decisions that will lead to profitable operations in the future. The profitability depends, on the one hand, on revenue-generating agreements with shippers that fix price and service levels and, on the other hand, on the cost of serving the shippers' customers. The strategic facility location decisions take place before procurement processes that fix the agreements and, consequently, before the operational planning of service to customers. The LP, therefore, has to take into account the agreements and their links to the cost of serving customers at a high level and under imperfect information. In this context, we adopt a highly abstract perspective on the complex organization of this industry. Whereas we recognize that shippers are profit-maximizing actors that procure services from the LP and then sell them to their customers, we do not explicitly model their respective profit maximization problem. Instead, we model shipper preferences

under the assumption that their sensitivity to price is implicitly dominated by their customers' preferences.

Agreements with shippers fix service levels and related prices. Specifically, the LP defines a subset $M_{nk} \subset M$ of offered service levels for each shipper and customer category (denoted by binary variables z_{nk}^m) as well as the associated prices t_n^m, for each $n \in N$ and $m \in M_{nk}$. Note that the prices t_n^m are not indexed by k as the LP does not control the pricing shippers define to their customers. The cost of allocating customer demand d_j, $j \in J$, to facility $i \in I$ at a variable cost c_{ij}^m depends on the service level $m \in M$ as well as on the geographical location of the demand and the facility. Associated with each facility is a fixed opening cost f_i and capacity u_i. We denote by r_i the binary decision variables related opened facilities, and by w_{ij}^m the decision variables related to the assignment of customers to facilities, which can be strict (binary) or not (in [0,1]) depending on the problem setting. Note that the service levels may lead to different capacity usage of the facilities. For example, a higher service level may require more resources.

Each shipper decides to accept or reject the price and service level alternatives offered by the LP. If the shipper accepts, then the demands from the corresponding customers are included in the agreement. Shippers can also opt to drop some of their customers from the agreement if, given the LP's offer, they prefer insourcing their own logistics operations or outsourcing them to other LPs. This option is henceforth referred to as *opt-out*. Consequently, having chosen a given shipper, the customers are captive to the shipper's decisions and face the service levels fixed in the agreement. Customers' preferences determine their choice of shipper and choice of service levels. Shippers, therefore, need to take their customers' preferences into account when concluding agreements with LPs. The decisions made by the shippers, of accepting or rejecting offers made by the LP, denoted by binary variables x_{nk}^m, hence depend on the shippers as well as on their customers' preferences. At this strategic decision level, the LP has imperfect information about such preferences but can have knowledge about certain customer attributes. Moreover, the LP makes facility location decisions before the shipper and customer preferences are revealed.

In summary, we consider a strategic decision-making problem faced by an LP seeking to open facilities that lead to profitable operations. When making such decisions, it takes into account the expected revenue generated by agreements, anticipating the reactions of the shippers, and the costs arising when satisfying those agreements (i.e., serving the shippers' customers). Given this hierarchical decision-making problem and conflicting objectives between the profit-maximizing LP and the shippers' cost-minimizing customers, the problem naturally lends itself to a bilevel formulation that we delineate in Section 2.4.

2.4. Mathematical Programming Formulation

This section expounds the mathematical formulation of our problem. Its underlying assumptions and a compact form of the formulation are outlined in Section 2.4.1. The latter clearly shows the stochastic bilevel structure of our problem. Section 2.4.2 presents the detailed constraints and a simulation-based version of the formulation.

2.4.1. Preliminaries

We introduce the assumptions and the compact formulation on which we build our subsequent developments. In our strategic setting, we consider the pricing problem at a high level and therefore assume the existence of a finite set of prices and service levels based on which the LP defines its offers to the shippers and their customers.

Assumption 1. *The LP decides among a finite set of discrete price levels* $\{q^p : p \in P\}$*, where* $q^1 < q^2 < \cdots < q^{|P|}$.

Assumption 2. *For each shipper* $n \in N$*, and each service level* $m \in M_n$*, there is a subset of possible price levels* $\{q_n^{mp} : p \in P_n^m \subseteq P\}$.

Based on this assumption, the price t_n^m charged for service level m to shipper n can only take discrete values. We introduce the binary decision variables y_n^{mp} taking value 1 if price q_n^{mp} is selected, and 0 otherwise. Hence, we have

$$t_n^m = \sum_{p \in P_n^m} q_n^{mp} y_n^{mp}. \tag{2.1}$$

Assumption 3. *Each service level m has a capacity usage proportional to a scale factor $\gamma^m \geq 1$.*

Assumption 4. *A minimum demand level l_n^{mp} is required from shipper n to have access to service level m at price level q_n^{mp}.*

Assumption 5. *Alternatives from competitors are grouped into one unique alternative service level, called opt-out, denoted by 0.*

Assumption 6. *$P_n^0 = \emptyset$, that is, there is no price available for opt-out alternative $m = 0$.*

Based on Assumption 6, customers in category k of a given shipper n can choose an alternative from the set $m \in M_{nk} \cup \{0\}$, which is simply denoted as M_{nk}^0.

We now introduce assumptions related to modeling shipper and customer preferences with RUM models. We first focus on the deterministic (observable) part of the utilities in Assumptions 7 and 8.

Assumption 7. *The preference of each shipper n for a service level m is modeled with a utility U_n^m that is additive in customer-specific utilities, U_{nk}^m, $k \in K_n$. That is, $U_n^m = \sum_{k \in K_n} U_{nk}^m$.*

Consequently, maximizing customer-specific utilities leads to maximizing the corresponding shipper's utility. Moreover, the price that the shipper accepts for the corresponding service level m affects the utility of each customer category, as stated in the following assumption.

Assumption 8. *The deterministic utility for a given price level is $U_{nk}^m = \alpha t_n^m + L_{nk}^m$ where α is a parameter capturing price sensitivity, and L_{nk}^m is an exogenous variable capturing customers' preferences for the service level.*

Using (2.1), we write utilities U_{nk}^m as

$$U_{nk}^m(\mathbf{y}) = \alpha\Bigg(\sum_{p\in P_n^m} q_n^{mp} y_n^{mp}\Bigg) + L_{nk}^m. \tag{2.2}$$

Moreover, based on Assumption 6, the utility of the opt-out alternative is $U_{nk}^0 = L_{nk}^0$ since $\sum_{p\in P_n^0} q_n^{0p} y_n^{0p} = 0$.

To simplify the mathematical expressions, we henceforth use a shorthand notation to designate vectors made up of individually indexed variables. Thus, $\mathbf{y}$ stands for the vector made up of $y_n^{mp}, \forall n, m, p$. Similarly, $\boldsymbol{\varepsilon}, \mathbf{z}, \mathbf{x}, \mathbf{r}, \mathbf{w}$ respectively stand for $\varepsilon_{nk}^m, z_{nk}^m, x_{nk}^m, \forall m,n,k$; $r_i, \forall i \in I$; $w_{ij}^m, \forall i, j, m$.

Given our strategic setting and the inherited uncertainty surrounding the demand distribution, we assume that utilities are not perfectly known. Adopting the additive RUM hypothesis, we assume that utilities include additive noise.

Assumption 9. $U_{nk}^m(\mathbf{y}; \varepsilon_{nk}^m) = \alpha\Big(\sum_{p\in P_n^m} q_n^{mp} y_n^{mp}\Big) + L_{nk}^m + \varepsilon_{nk}^m$, *where* ε_{nk}^m *is a continuous random variable with support* Ξ.

Note in Assumption 9 we distinguish in the notation, the deterministic (decision vector $\mathbf{y}$) from the stochastic (ε_{nk}^m) inputs, by separating them with a semicolon. Finally, we make an assumption about the assignment of customers to facilities reflecting the practice that multiple facilities can service a given customer.

Assumption 10. *The assignment of customers to facilities is not strict, that is,* $0 \leq w_{ij}^m \leq 1$.

Our problem can naturally be formalized as a bilevel program where the LP and the shippers respectively play the roles of the leader and the followers. To clearly show the structure of our problem, we introduce a compact form

$$\max \mathbf{E}_{\boldsymbol{\varepsilon}}[G(\mathbf{y},\mathbf{z},\mathbf{r},\mathbf{x},\mathbf{w}) : (\mathbf{y},\mathbf{z},\mathbf{r},\mathbf{w}) \in H(\mathbf{x}); \mathbf{x} \in F(\mathbf{y},\mathbf{z};\boldsymbol{\varepsilon})]. \tag{2.3}$$

Although we have not provided details about the RUM model yet, stochasticity is implicit. The LP's objective function $G(\cdot)$ depends on decision variables $\mathbf{y}$,$\mathbf{z}$,$\mathbf{r}$,$\mathbf{x}$,$\mathbf{w}$, and constraints $H(\mathbf{x})$ depend on the followers' (i.e., shippers') decision variables $\mathbf{x}$.

In turn, optimal decisions of shippers $\mathbf{x} \in F(\mathbf{y},\mathbf{z};\boldsymbol{\varepsilon})$ depend on the LP's decision variables $\mathbf{y}$ and $\mathbf{z}$ as well as on random variables $\boldsymbol{\varepsilon}$. Two remarks are in order. First, two vectors of the leader's decision variables – $\mathbf{r}$ and $\mathbf{w}$ – do not impact the followers' problem. Second, the objective function $G(\cdot)$ as well as the set $H(\mathbf{x})$ are implicitly affected by the random variables $\boldsymbol{\varepsilon}$ through shippers' decisions $\mathbf{x}$. This is one of the challenging aspects that we focus on in subsequent sections.

The set of optimal solutions of the shippers is

$$F(\mathbf{y},\mathbf{z};\boldsymbol{\varepsilon}) = \arg\max_{\mathbf{x}} \left[\sum_{n\in N} \sum_{k\in K_n} \sum_{m\in M^0_{nk}} U^m_{nk}(\mathbf{y},\varepsilon^m_{nk}) x^m_{nk} : \mathbf{x} \in R(\mathbf{z}) \right]. \tag{2.4}$$

We note that in our context the solution is unique, as we discuss when introducing Proposition 1 in Section 2.5.1. The feasible set of shippers' solutions for given LP's decisions is

$$R(\mathbf{z}) = \left\{ \sum_{m\in M^0_{nk}} x^m_{nk} = 1, \forall n \in N, k \in K_n \; ; \; x^m_{nk} \leq z^m_{nk}, \forall n \in N, k \in K_n, m \in M_{nk} \right\}. \tag{2.5}$$

The first set of constraints imposes that the shipper can choose exactly one of the service levels proposed by the LP, or the opt-out alternative. The second set of constraints requires that the shipper can choose the service level only if it was offered to the shipper and the customer category. The random variables $\boldsymbol{\varepsilon}$ affect the objective function of the shippers but not the set $R(\mathbf{z})$. The shipper's objective is to maximize utility $U^m_{nk}(\mathbf{y};\boldsymbol{\varepsilon})$, and, assuming a negative perception of price ($\alpha < 0$), this objective is conflicting with the profit-maximizing objective of the LP. The demand allocation decisions $\mathbf{w}$ do not affect the set of optimal decisions $F(\mathbf{y},\mathbf{z};\boldsymbol{\varepsilon})$ as they do not impact the service perceived by the shippers. Notice that the shipper's problem can be decomposed: The set $F(\mathbf{y},\mathbf{z};\boldsymbol{\varepsilon})$ can be partitioned as $F(\mathbf{y},\mathbf{z};\boldsymbol{\varepsilon}) = \bigcup_{n\in N,k\in K_n} F_{nk}(\mathbf{y},\mathbf{z};\boldsymbol{\varepsilon})$ where $F_{nk}(\mathbf{y},\mathbf{z};\boldsymbol{\varepsilon})$ represents the set of optimal solutions for each shipper n and customer category k.

2.4.2. Simulation-based Formulation

This section presents a formulation based on the compact form (2.3) where the constraints and the RUM model are described in detail.

Choice probabilities arising from RUM models are non-linear in the decision variables. This is one of the challenges associated with their use in mathematical programming formulations. Essentially, this has been dealt with in two different ways in the literature: either through linearization of the well-known logit model (e.g. Haase and Müller, 2014), or by using simulated linear utilities as opposed to probabilities, and relying on sample average approximation (SAA) (Pacheco Paneque et al., 2021). We adopt the latter approach as it allows to use any RUM model as long as it is possible to simulate utilities. Accordingly, in this section we introduce the SAA corresponding to (2.3). Since the shippers' preferences are revealed only after the LP's strategic decisions, we use a two-stage stochastic bilevel structure. Our first-stage decision variables are the prices y_n^{mp}, the service level assignments z_{nk}^m, and the facility location decisions r_i. The LP can only allocate demand to facilities once shippers accept to include those customers in the agreements. Therefore, the second-stage decision variables are shippers' binary decisions to accept or reject service level offers, x_{nk}^m, and the LP's demand allocation decisions, w_{ij}^m.

Our first-stage problem is

$$\max_{\mathbf{r},\mathbf{z},\mathbf{y}} -\sum_{i\in I} f_i r_i + Q(\mathbf{r},\mathbf{z},\mathbf{y}) \tag{2.6a}$$

$$\text{s.t.} \sum_{p\in P_n^m} y_n^{mp} \leq 1, \qquad n\in N, m\in M_n, \tag{2.6b}$$

$$\sum_{m\in M_n}\sum_{p\in P_n^m} y_n^{mp} \leq |K_n|, \qquad n\in N, \tag{2.6c}$$

$$\sum_{m\in M_{nk}} z_{nk}^m \leq 1, \qquad n\in N, k\in K_n, \tag{2.6d}$$

$$z_{nk}^m \leq \sum_{p\in P_n^m} y_n^{mp}, \qquad n\in N, k\in K_n, m\in M_n, \tag{2.6e}$$

$$r_i \in \{0,1\}, \qquad i \in I, \tag{2.6f}$$

$$y_n^{mp} \in \{0,1\}, \qquad n \in N, m \in M_n, p \in P_n^m, \tag{2.6g}$$

$$z_{nk}^m \in \{0,1\}, \qquad n \in N, k \in K_n, m \in M_n. \tag{2.6h}$$

The objective function (2.6a) is the expected profit. Its first term captures the fixed cost of opening facilities and the second term is equal to the expected value of the second-stage problem $Q(\mathbf{r},\mathbf{z},\mathbf{y}) = \mathbf{E}_{\boldsymbol{\varepsilon}}\left[\phi(\mathbf{r},\mathbf{z},\mathbf{y};\boldsymbol{\varepsilon})\right]$. Constraints (2.6b) impose that the LP can choose at most one price for each shipper and service level. Constraints (2.6c) ensure that the maximum number of price and service level combinations assigned to a shipper does not exceed the number of its customer categories. Constraints (2.6d) impose that the LP can choose at most one service level for each shipper and customer category. Constraints (2.6e) ensure that a customer category can be assigned to a service level only if the service level has been associated to a price level.

Next, we turn our attention to the second-stage problem which is a bilevel program. Using the approach proposed in Pacheco Paneque et al. (2021), we estimate $Q(\mathbf{r},\mathbf{z},\mathbf{y})$ with an SAA,

$$\bar{Q}(\mathbf{r},\mathbf{z},\mathbf{y}) = \frac{1}{|S|}\sum_{s\in S}\phi_s(\mathbf{r},\mathbf{z},\mathbf{y};\boldsymbol{\varepsilon}_s), \tag{2.7}$$

where $\boldsymbol{\varepsilon}_s$ is a realization of $\boldsymbol{\varepsilon}$ for scenario $s \in S$ (S is the set of scenarios drawn independently at random). For a given s, the second-stage, upper-level problem is

$$\phi_s(\mathbf{r},\mathbf{z},\mathbf{y};\boldsymbol{\varepsilon}_s) = \max_{\mathbf{w}} \sum_{n\in N}\sum_{k\in K_n}\sum_{m\in M_k}\sum_{p\in P_n^m} d_k q_n^{mp} x_{nks}^m y_n^{mp} - \sum_{i\in I}\sum_{j\in J}\sum_{m\in M_j} c_{ij}^m d_j w_{ijs}^m \tag{2.8a}$$

$$\text{s.t.} \sum_{j\in J}\sum_{m\in M_j} w_{ijs}^m \gamma^m d_j \le u_i r_i, \qquad i \in I, \tag{2.8b}$$

$$\sum_{m\in M_j} w_{ijs}^m \le r_i, \qquad i \in I, j \in J_i, \tag{2.8c}$$

$$\sum_{i\in I} w_{ijs}^m = x_{nks}^m, \qquad n \in N, k \in K_n, j \in J_k, m \in M_j, \tag{2.8d}$$

$$\sum_{p \in P_n^m} l_n^{mp} y_n^{mp} \leq \sum_{k \in K_n} d_k x_{nks}^m, \qquad n \in N, m \in M_n, \tag{2.8e}$$

$$x_{nks}^m \in F_{nks}(\mathbf{y}, \mathbf{z}; \boldsymbol{\varepsilon}_s), \qquad n \in N, k \in K_n, m \in M_n, \tag{2.8f}$$

$$0 \leq w_{ijs}^m \leq 1, \qquad i \in I, j \in J, m \in M_j. \tag{2.8g}$$

The first term in the objective function in (2.8a) corresponds to the revenue generated in this scenario, and the second term to the variable costs. Note that the latter depends on the assignment of specific customers to facilities. Constraints (2.8b) ensure that the demands assigned to facilities do not exceed the respective capacities. Constraints (2.8c) impose that customers are assigned to open facilities only. Constraints (2.8d) require that only the demand of customers included in the agreements is allocated to any facility. Constraints (2.8e) ensure that the required minimum demand level for price p and service level m is respected.

In Constraints (2.8f), the set of optimal solutions for the followers' (lower-level) problem corresponding to a shipper and customer category is

$$F_{nks}(\mathbf{y},\mathbf{z};\boldsymbol{\varepsilon_s}) = \arg\max_{\mathbf{x}} \left[\sum_{m \in M_{nk}^0} U_{nks}^m(\mathbf{y}; \varepsilon_{nks}^m) x_{nks}^m : \sum_{m \in M_{nk}^0} x_{nks}^m = 1; x_{nks}^m \leq z_{nk}^m; x_{nks}^m \in \{0,1\} \right], \tag{2.9}$$

and it depends on the leader's decisions $\mathbf{y}$ and $\mathbf{z}$. Note that the simulated utility $U_{nks}^m(\mathbf{y}; \varepsilon_{nks}^m) = \alpha\left(\sum_{p \in P_n^m} q_n^{mp} y_n^{mp} \right) + L_{nk}^m + \varepsilon_{nks}^m$ is linear in decision variables whereas the objective function is bilinear. Moreover, when $\alpha < 0$, the utility decreases as price increases, in contrast to (2.8a), which increases with price.

This two-stage stochastic bilevel formulation cannot be solved with a general-purpose MILP solver. In the following section, we describe the solution approaches used in this work.

2.5. Solution Approaches

This section describes exact and heuristic solution approaches for the SAA bilevel model introduced in Section 2.4. First, in Section 2.5.1, we show that we can compute optimal solutions to the follower's problem (2.9) by solving a continuous unconstrained relaxation. The overall problem can then be solved by a standard Benders

decomposition method. Then, in Section 2.5.2, we introduce our heuristic that is based on a preprocessing phase and solving a reduced single-level MILP formulation.

2.5.1. Continuous Relaxation of the Follower's Problem

The following proposition establishes the foundation for a single-level MILP model.

Proposition 1. *Optimal solutions to the follower's problem* (2.9) *can be computed by solving, for a given* $m' \in M_{nk}$*, a continuous relaxation*

$$\arg\max \left[\left(U^{m'}_{nks}(\mathbf{y}; \varepsilon^{m'}_{nks}) - U^{0}_{nks}(\varepsilon^{0}_{nks}) \right) x^{m'}_{nks} : x^{m'}_{nks} \in [0,1] \right]. \tag{2.10}$$

Proof. By replacing $x^0_{nks} = 1 - \sum_{m \in M_{nk}} x^m_{nks}$ in (2.9) and rearranging the corresponding terms $x^m_{nks}, m \in M_{nk}$, we obtain

$$\arg\max \left[\sum_{m \in M_{nk}} \left(U^{m}_{nks}(\mathbf{y}; \varepsilon^{m}_{nks}) - U^{0}_{nks}(\varepsilon^{0}_{nks}) \right) x^{m}_{nks} : x^{m}_{nks} \leq z^{m}_{nk}; x^{m}_{nks} \in \{0,1\}, m \in M_{nk} \right]. \tag{2.11}$$

Recall that (2.11) is solved for fixed z^m_{nk} and y^{mp}_n. Due to Constraints (2.6d), either $z^m_{nk} = 0, \forall m \in M_{nk}$ and as a consequence $x^m_{nks} = 0$, or we solve the problem for a given m' where $z^{m'}_{nk} = 1$. That is,

$$\arg\max \left[\left(U^{m'}_{nks}(\mathbf{y}; \varepsilon^{m}_{nks}) - U^{0}_{nks}(\varepsilon^{0}_{nks}) \right) x^{m'}_{nks} : x^{m'}_{nks} \leq 1; x^{m'}_{nks} \in \{0,1\} \right]. \tag{2.12}$$

If we relax the binary restriction in (2.12), that is $0 \leq x^{m'}_{nks} \leq 1$, three cases can be distinguished in regard to the objective function in (2.9). First, $U^{m'}_{nks}(\mathbf{y}; \varepsilon^{m'}_{nks}) - U^{0}_{nks}(\varepsilon^{0}_{nks}) > 0$, in which case $x^{m'}_{nks} = 1$ is optimal. Second, $U^{m'}_{nks}(\mathbf{y}; \varepsilon^{m'}_{nks}) - U^{0}_{nks}(\varepsilon^{0}_{nks}) < 0$, in which case $x^{m'}_{nks} = 0$ is optimal. Third, $U^{m'}_{nks}(\mathbf{y}; \varepsilon^{m'}_{nks}) - U^{0}_{nks}(\varepsilon^{0}_{nks}) = 0$. However, the probability that this occurs is zero as the samples are drawn from continuous distributions. Hence, solving (2.10) yields the optimal solution to (2.9). □

Proposition 1 shows that the follower's problem is a linear program. Using the strong duality theorem and standard linearization techniques, we reformulate

(2.8) as a single-level MILP model. The corresponding reformulation is shown in Appendix A.2. We provide a few related notes. First, since the probability that $U^{m'}_{nks}(\mathbf{y};\varepsilon^{m'}_{nks}) - U^{0}_{nks}(\varepsilon^{0}_{nks}) = 0$ is zero, the solution to the follower problem (2.9) is unique for a given $\mathbf{y}$, $\mathbf{z}$ and ε. Second, for fixed $\mathbf{y}$, $\mathbf{z}$ and $\mathbf{r}$, the resulting single-level reformulation of the second-stage problem (2.8) is also a linear program. Third, the reformulation involves so-called big-M constants. Note that *bilevel-correct* big-M constants can be obtained efficiently, as stated in Buchheim (2023). However, computing the smallest one is still challenging in the general case when the single-level reformulation is obtained using the Karush–Kuhn–Tucker conditions (e.g., Kleinert et al., 2020). We obtain the single-level reformulation using strong duality. Nevertheless, big-M constants arise when we linearize bilinear terms. We provide problem-specific bounds for these constants that are valid for any leader decision (see Proposition 6 in Appendix A.2). Hence, a standard Benders decomposition method can be applied, with master problem defined by (2.6a) - (2.6h) and $|S|$ subproblems to compute $\phi_s(\mathbf{r},\mathbf{z},\mathbf{y};\varepsilon_{\mathbf{s}})$.

2.5.2. Single-level Reduced Formulation and Heuristic

In this section, we introduce a heuristic solution approach. The main idea is to solve a formulation that has a reduced size because we only consider scenarios having a non-zero effect on the objective function. We, therefore, start with a result that allows us to identify first-stage solutions leading to second-stage decision variables of zero value. We refer to them as *trivial solutions*.

Proposition 2. *For a given $s \in S$, and for given n,k,m and p with $z^m_{nk} = 1$ and $y^{mp}_n = 1$ leading to $U^m_{nks}(y^{mp}_n;\varepsilon^m_{nks}) < U^0_{nks}(\varepsilon^0_{nks})$, we have*

$$\left(U^m_{nks}(y^{mp}_n;\varepsilon^m_{nks}) - U^0_{nks}(\varepsilon^0_{nks})\right) x^m_{nks} = 0,$$

$$d_k q^{mp}_n x^m_{nks} y^{mp}_n - \sum_{i\in I} \sum_{j\in J'_n \cap J_k} c^m_{ij} d_j w^m_{ijs} = 0.$$

Proof. If $U^m_{nks}(y^{mp}_n;\varepsilon^m_{nks}) < U^0_{nks}(\varepsilon^0_{nks})$ then, as explained in the proof of Proposition 1, $x^m_{nks} = 0$. Therefore, due to Constraints (2.8d), $w^m_{ijs} = 0, \forall j \in J^{'}_n \cap J_k, i \in I$. □

We partition the set of scenarios S, distinguishing those that lead to trivial solutions from those that do not. Specifically, for each n, k, m and p, $S^{mp}_{nk} = \{s \in S : U^m_{nks}(y^{mp}_n;\varepsilon^m_{nks}) - U^0_{nks}(\varepsilon^0_{nks}) > 0\}$, and $(S^{mp}_{nk})^c = \{s \in S : U^m_{nks}(y^{mp}_n;\varepsilon^m_{nks}) - U^0_{nks}(\varepsilon^0_{nks}) \leq 0\}$. For each z^m_{nk} and y^{mp}_n assignment, shipper n accepts the offer for customers in category k, $x^m_{nks} = 1$, in a share $|S^{mp}_{nk}|/|S|$ of the scenarios.

This share $|S^{mp}_{nk}|/|S|$ is the SAA of the probability

$$\rho^{mp}_{nk} = P\left(U^m_{nk}(y^{mp}_n = 1;\varepsilon^m_{nk}) > U^0_{nk}(\varepsilon^0_{nk}) | z^m_{nk} = 1\right) = P\left(\alpha q^{mp}_n + L^m_{nk} + \varepsilon^m_{nk} > L^0_{nk} + \varepsilon^0_{nk}\right), \tag{2.13}$$

and we have

$$\lim_{|S|\to\infty} \frac{|S^{mp}_{nk}|}{|S|} \to \rho^{mp}_{nk}. \tag{2.14}$$

Probabilities ρ^{mp}_{nk} play a central role in our heuristic approach that is based on precomputing these probabilities and then solving a reduced formulation that does not explicitly depend on the scenarios. In the following, we describe how we derive the reduced formulation, and then we discuss how we compute the probabilities ρ^{mp}_{nk}.

To introduce ρ^{mp}_{nk}, we rewrite the objective function (2.8a) of the leader in the second-stage problem:

$$\max \sum_{n\in N}\sum_{k\in K_n}\sum_{m\in M_k}\sum_{p\in P^m_n} d_k q^{mp}_n x^m_{nks} y^{mp}_n - \sum_{n\in N}\sum_{k\in K_n}\sum_{i\in I}\sum_{j\in J'_n\cap J_k}\sum_{m\in M_j} c^m_{ij} d_j w^m_{ijs} \sum_{p\in P^m_n} y^{mp}_n. \tag{2.15}$$

In brief, the problem then consists in finding the combinations of prices and service levels with $\rho^{mp}_{nk} > 0$ that provide the maximum profit to the LP. Rewriting the objective in this way is straightforward for the revenue terms, whereas the cost terms require introducing variables y^{mp}_n (see Proposition 7 in Appendix A.4).

When a service level and price alternative (m,p) assigned to a shipper n leads to a negative profit, the optimal solution for the leader is $z^m_{nk} = 0$ and $y^{mp}_n = 0$, and hence, $x^m_{nks} = 0$, and $w^m_{ijs} = 0$. This implies that the impact on the objective function is zero for all of the corresponding scenarios. On the contrary, when $z^m_{nk} = 1$, and

$y_n^{mp} = 1$, for (m,p), it maximizes the utility of shipper n and customers in category k for a fraction $|S_{nk}^{mp}|/|S|$ of the scenarios. Therefore, for each n, k, m and p, the corresponding scenarios $s \in (S_{nk}^{mp})^c$ in the objective function can be discarded, as they do not provide an optimal solution for the followers (the shippers) nor a profit to the LP. In addition, because we are only considering the scenarios in which the alternatives proposed by the leader maximize the utility of the followers, the decision variables x_{nks}^m can be replaced by the corresponding z_{nk}^m as shown in the following proposition.

Proposition 3. *For a given assignment z_{nk}^m and y_n^{mp}, $x_{nks}^m = z_{nk}^m$ for all $s \in S_{nk}^{mp}$.*

Proof. Recall $x_{nks}^m \leq z_{nk}^m$ as seen in (2.9). If $z_{nk}^m = 0$, then $x_{nks}^m = 0$, for all $s \in S_{nk}^{mp}$. If $z_{nk}^m = 1$, then $x_{nks}^m = 1$ for all $s \in S_{nk}^{mp}$, as $U_{nks}^{mp} > U_{nks}^0$ for $s \in S_{nk}^{mp}$, otherwise the solution would not be optimal for the follower and would therefore violate (2.8f). Hence, $x_{nks}^m = z_{nk}^m$ for $s \in S_{nk}^{mp}$. □

This leads us to the following reduced single-level formulation that is an approximation of (2.7):

$$\max \sum_{n\in N}\sum_{k\in K_n}\sum_{m\in M_k}\sum_{p\in P_n^m} \frac{|S_{nk}^{mp}|}{|S|} d_k q_n^{mp} z_{nk}^m y_n^{mp} - \sum_{n\in N}\sum_{k\in K_n}\sum_{i\in I}\sum_{j\in J_n'\cap J_k}\sum_{m\in M_{nk}}\sum_{p\in P_n^m} \frac{|S_{nk}^{mp}|}{|S|} c_{ij}^m d_j w_{ij}^m y_n^{mp} \tag{2.16a}$$

$$\text{s.t.} \sum_{j\in J}\sum_{m\in M_j} w_{ij}^m \gamma^m d_j \leq u_i r_i, \qquad i \in I, \tag{2.16b}$$

$$\sum_{m\in M_j} w_{ij}^m \leq r_i, \qquad i \in I, j \in J_i, \tag{2.16c}$$

$$\sum_{i\in I} w_{ij}^m = z_{nk}^m, \qquad n \in N, k \in K_n, j \in J_k, m \in M_j, \tag{2.16d}$$

$$\sum_{k\in K_n} d_k z_{nk}^m \geq \sum_{p\in P_n^m} l_n^{mp} y_n^{mp}, \qquad n \in N, m \in M_n, \tag{2.16e}$$

$$0 \leq w_{ij}^m \leq 1, \qquad i \in I, j \in J, m \in M_j. \tag{2.16f}$$

Note that there are bilinear terms in (2.16a) the objective. We linearize them by introducing a new set of variables $\pi_{nk}^{mp} :\equiv y_n^{mp} z_{nk}^m$, $\nu_{nk}^m :\equiv w_{ij}^m y_n^{mp}$:

$$\pi_{nk}^{mp} \leq z_{nk}^m, \qquad n \in N, k \in K_n, m \in M_{nk}, p \in P_n^m, \tag{2.17a}$$

$$\pi_{nk}^{mp} \leq y_n^{mp}, \qquad n \in N, k \in K_n, m \in M_{nk}, p \in P_n^m, \tag{2.17b}$$

$$\pi_{nk}^{mp} \geq z_{nk}^m + y_n^{mp} - 1, \qquad n \in N, k \in K_n, m \in M_{nk}, p \in P_n^m, \tag{2.17c}$$

$$\pi_{nk}^{mp} \geq 0, \qquad n \in N, k \in K_n, m \in M_{nk}, p \in P_n^m, \tag{2.17d}$$

$$\nu_{ij}^{mp} \leq w_{ij}^m, \qquad n \in N, j \in J_n^{'}, i \in I, m \in M_j, \tag{2.17e}$$

$$\nu_{ij}^{mp} \leq y_n^{mp}, \qquad n \in N, j \in J_n^{'}, i \in I, m \in M_j, \tag{2.17f}$$

$$\nu_{ij}^{mp} \geq w_{ij}^m + y_n^{mp} - 1, \qquad n \in N, j \in J_n^{'}, i \in I, m \in M_j, \tag{2.17g}$$

$$\nu_{ij}^{mp} \geq 0, \qquad n \in N, j \in J_n^{'}, i \in I, m \in M_j. \tag{2.17h}$$

Finally, we obtain the reduced single-level formulation, an approximation of (2.6),

$$\begin{aligned}
\max & -\sum_{i \in I} f_i r_i + \sum_{n \in N} \sum_{k \in K_n} \sum_{m \in M_k} \sum_{p \in P_n^m} \frac{|S_{nk}^{mp}|}{|S|} d_k q_n^{mp} \pi_{nk}^{mp} \\
& - \sum_{n \in N} \sum_{k \in K_n} \sum_{i \in I} \sum_{j \in J_n' \cap J_k} \sum_{m \in M_{nk}} \sum_{p \in P_n^m} \frac{|S_{nk}^{mp}|}{|S|} c_{ij}^m d_j \nu_{ij}^{mp} \\
\text{s.t. } & (2.6b) - (2.6h), (2.16b) - (2.16f), (2.17a) - (2.17h),
\end{aligned}$$

that we solve as part of our heuristic. Note that, constraints (2.6b) - (2.6h) remain unchanged as they correspond to the first-stage problem and, therefore, are not affected by the scenarios.

Before solving the formulation, we compute ρ_{nk}^{mp} (using (2.13) and (2.14)) in a preprocessing phase. More precisely, they can be computed by simulation (see Algorithm 1 where the function RandomUtilityGenerator is tailored to the specific RUM), or through a closed-form, if it exists.

To verify the accuracy of our proposed heuristic, we apply in Section 2.7 a standard Benders decomposition method to solve the original formulation introduced in Section 2.5.1. There, we consider two variants of the heuristic: analytical or scenario-based, depending on how we compute ρ_{nk}^{mp} in the preprocessing phase.

Algorithm 1 Computing ρ_{nk}^{mp} by simulation

Require: $q_n^{mp}, L_{nk}^m, L_{nk}^0, S$
  sum $\leftarrow 0$
  for all $s \in S$ **do**
    $U_{nks}^{mp} \leftarrow$ RandomUtilityGenerator(q_n^{mp},L_{nk}^m)
    $U_{nks}^0 \leftarrow$ RandomUtilityGenerator(L_{nk}^0)
    if $\left(U_{nks}^{mp} > U_{nks}^0\right)$ **then**
      sum $\leftarrow$ sum $+ 1$
    end if
  end for
  $\rho_{nk}^{mp} \leftarrow$ sum$/|S|$

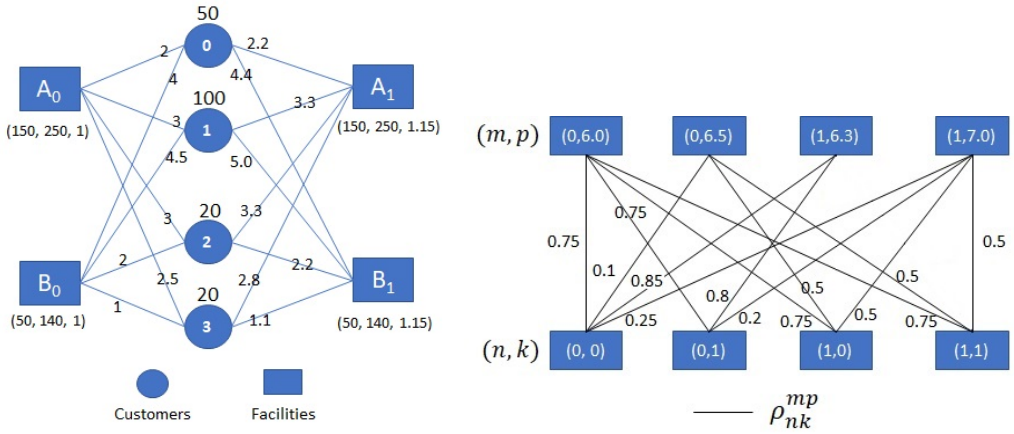
2.6. Illustrative Example

This section introduces a small example to illustrate the impact of the RUM. To this aim, we consider three cases: The first case features a particular distribution over offered alternatives and the opt-out. The second case is a perfect information version in which the opt-out is not selected. The third case is a maximum uncertainty version where it is equally likely that the shippers select the opt-out as one of the offered alternatives. Before summarizing the findings for each of these cases, we describe the core settings of the example.

There are two shippers, $N = \{0,1\}$, each having a set of two customers: $J_0^{'} = \{0,1\}$ and $J_1^{'} = \{2,3\}$. The demand levels for each customer of Shipper 0 are 50 and 100, whereas they are 20 for both customers of Shipper 1. Hence, the total demand from Shipper 0 is 150, and it is 40 for Shipper 1. The LP has the option to open two facilities, $I = \{A,B\}$, with a capacity of 150 and 50, respectively. The fixed cost of opening facility A is 250, and that of facility B is 140. We depict the structure of the facility location part of the problem in Figure 2.1(a). We show the values (u_i,f_i,γ_m) below each facility node i. Note that we illustrate the two facilities – A and B – twice to show that the assignment costs c_{ij}^m are affected by the service levels. We use sub-indexes when denoting the facilities to identify the corresponding service level. We also report the demand of each customer above the corresponding node.

The LP offers the same two service levels to both shippers, $M = \{0,1\}$, with capacity usage rate of $\gamma^1 = 1$ and $\gamma^2 = 1.15$. Furthermore, they are both offered with the same price levels, $P^0 = \{6.0, 6.5\}$ and $P^1 = \{6.3, 7.0\}$, for services 0 and 1, respectively. However, the lower prices are only available if a minimum demand level is satisfied: 40 and 50 units for levels 0 and 1, respectively (i.e., $l_0^{0,6.0} = l_1^{0,6.0} = 40$ and $l_0^{1,6.3} = l_1^{1,6.3} = 50$). Since the total demand for Shipper 1 is 40, it cannot access the cheaper price for service level 1.

In the first case, the probabilities, shown next to the arcs in Figure 2.1(b), are fixed such that they are higher for lower prices than for higher prices. Moreover, it is more likely that the opt-out will be selected when prices are higher. Note that the sum of the probabilities of selecting an alternative, and the opt-out (not shown in the figure) equals one.



(a) Network structure **(b)** Probabilities according to the RUM

Fig. 2.1. Overview of the illustrative example setup

In this setting, the optimal solution with an expected profit of 125 consists in opening Facility A and serve all demand from Shipper 0 at Service level 0 and at price 6.0. Indeed, Shipper 0 then meets the minimum demand level. It is not optimal to serve Shipper 1. In this case Facility B would be required and it is not profitable as the expected revenue is less than the fixed cost of opening that facility. Note that

only Service level 0 can be assigned to customers of Shipper 0. In case Service level 1 would be chosen, the demand would exceed the capacity of Facility A ($1.5 \cdot 150 > 150$).

Turning to the second case, we assume perfect information and that the opt-out is not selected. This yields an upper bound on the expected profit. In this example it is 324, almost 3 times larger than the one discussed previously. In contrast, asssuming as a third case a uniform distribution ($\rho_{nk}^{mp} = 0.5; \forall n,k,m,p$) results in zero expected profit. Under such high uncertainty, it is not worth opening any of the facilities.

This small example illustrates that assumptions on shipper and customer preferences can have an important impact on expected profit and should, therefore, be taken into account when making strategic facility location decisions. In the following section, we report computational results and analyze a variety of instances and parameter settings.

2.7. Computational Results

This section reports on an extensive computational study. First, Section 2.7.1 describes the experimental setup in detail. Then, results are divided into two parts: Section 2.7.2 compares the solution quality obtained with the heuristic to optimal solutions computed with the Benders decomposition method. Section 2.7.3 assesses the impact of key parameters on the solution quality and computing times of the heuristic.

2.7.1. Experimental Setup

Since we introduce a new problem, we do not have access to existing benchmark instances and, as explained in Section 2.2, we are unaware of existing solution approaches that we could use as baselines. However, we can use the *capacitated facility location problem* (CFLP) structure of our problem to generate instances that are close to those in the literature. We use two existing sets of benchmark instances for this purpose. First, we generate test instances with the generator proposed by Klose (2018) and utilized in Klose and Görtz (2007) and Görtz and Klose (2011). Second, we use the set of CFLP instances found in the *OR-Library* of Beasley (1990).

We proceed to delineate the main characteristics of our experimental setup. Since the setup varies depending on the objective of the experiments, sections are divided accordingly. Hence, sections and their focus are as follows: (i) Section 2.7.2: solution quality compared to exact baseline, (ii) Section 2.7.3.1: impact of demand model parameters, (iii) Section 2.7.3.2: impact of CFLP parameters, and (iv) Section 2.7.3.3: impact of problem size (for the sake of readability, we omit some details about Section 2.7.3.3).

Logistics provider. In Section 2.7.2, we fix the numbers of facility locations $|I| = 2$, service levels $|M| = 2$ and price levels $|P| = 5$. We use $P = \{10, 12.5, 15, 17.5, 20\}$. Note that we intentionally keep the size of the instances small so that they can be solved to optimality even with a large number of scenarios. In Sections 2.7.3.1 and 2.7.3.2, we use $|I| = 4$, service levels $|M| = 3$ and price levels $|P| = 5$. We use 15 and 23 as lower and upper bounds for price levels, and equal price steps are given by $\frac{23-15}{|P|-1}$. For $|P| = 5$, we therefore set $P = \{15,17,19,21,23\}$. In Section 2.7.3.3, we vary $|I|$ and $|P|$ and we use $|P| \in \{3,4,5,10,15\}$, depending on the class of instances tested. For two of the classes of instances we test in that section, we generate the set of price levels as before.

For the sake of simplicity, we assume that the service levels do not impact capacity utilization (i.e., $\gamma^m = \gamma = 1$, $\forall m \in M$) in any of the results. We fix the relative costs of assigning customers to facilities for the three different service levels to be $c^1_{ij} = c_{ij}$, $c^2_{ij} = 1.05c_{ij}$ and $c^3_{ij} = 1.1c_{ij}$. Note that c_{ij}, along with facility capacities u_i, fixed opening costs f_i, and customer demands d_j are fixed by the CFLP instance generator, as we detail next.

CFLP generator. We use the instance generator of Klose (2018) with a random seed equal to 963490972. The values of c_{ij}, u_i, f_i and d_j are generated such that they satisfy a ratio $R = \frac{\sum_{i \in I} u_i}{\sum_{j \in J} d_j}$ relating the capacity of the facilities and the total demand of the customers. In Section 2.7.2, we fix this ratio to 1. In Section 2.7.3.1, we use 2 as the ratio, whereas we allow it to vary in Section 2.7.3.2. Finally, in Section 2.7.3.3, we use $R = \{5,10\}$ to derive easy and hard instances. In this section, we also use instances in *OR-Library* (Beasley, 1990) of the capacitated warehouse location problem. Thus, we adapted the following instances to our problem: Problems IV,

VI, VIII, IX, X, XI, XII, XII, and problem set A from the capacitated warehouse location problem test data sets (used in Beasley, 1988). We refer to them as *ORLib* instances. Note that instances in this class do not require a ratio R.

Shippers and customers. In Section 2.7.2, we fix the numbers of shippers $|N| = 1$, customers $|J| \in \{15{,}30{,}45\}$ and customer categories $|K| = 3$. In Sections 2.7.3.1 and 2.7.3.2, we use $|N| = 2$, $|J| = 48$ and $|K| = 3$. In Section 2.7.3.3, we use larger J sets and $|K| \in \{1{,}3\}$, depending on the class of instance used.

RUM model. In all experiments, we consider a logit model, i.e., ε are independently and identically distributed Extreme Value type I distribution with scale parameter β. For the sake of simplicity, we assume that all shipper and customer categories share a common utility specification

$$U_{nk}^{mp} = U^{mp} = \alpha_n^m q^{mp} + L_{nk}^m + \varepsilon = \alpha q^{mp} + 4.5 + \varepsilon. \tag{2.18}$$

In Section 2.7.2, the utility of the opt-out alternative U^0 takes one of the values in $\{0{,}4.5\}$. We set these values to analyze two cases, one where the service levels proposed by the LP are much better than the opt-out, and another where both are perceived identically by the shippers. In all subsections of Section 2.7.3, $U^0 = 3$ so that the deterministic part of the utility for the opt-out and that of price $q^0 = 15$ are equal. The value of parameter α represents the customers' price sensitivity. We fix $\alpha = -0.15$ in Section 2.7.2. In Section 2.7.3.2, we set $\alpha = -0.10$, whereas in Section 2.7.3.3 we use two values, $\alpha = -0.10$ and $\alpha = -0.15$, to define easy and difficult instances respectively. In Section 2.7.3.1, we vary the value of α (see Appendix A.3 for details). Regarding the parameter β, in Sections 2.7.2 and 2.7.3.3, we use 11 different values for β, as detailed in Appendix A.3. In Section 2.7.3.2, we fix $\beta = 1$, whereas in Section 2.7.3.3 we use $\beta \in \{0.5, 1\}$, to distinguish easy instances from difficult ones.

We compute the probabilities ρ_{nk}^{mp} in a preprocessing phase based on the utility model previously introduced. Given that we assume a logit model, we do not need to simulate the utilities as we can compute the theoretical probabilities. Moreover, as shippers and customers share a common utility specification, the probability is

computed as follows:

$$\rho_{nk}^{mp} = \rho^{mp} = \frac{\mathrm{e}^{\frac{1}{\beta}(-0{,}1q^{mp}+4{,}5)}}{\mathrm{e}^{\frac{1}{\beta}(-0{,}1q^{mp}+4{,}5)} + \mathrm{e}^{\frac{1}{\beta}3}} = \frac{\mathrm{e}^{\frac{1}{\beta}(-0{,}1q^{mp}+1{,}5)}}{\mathrm{e}^{\frac{1}{\beta}(-0{,}1q^{mp}+1{,}5)} + 1}.$$

Note, however, that we use simulated probabilities in Section 2.7.2 to allow for a fair comparison with the exact baseline.

Performance metrics. We analyze expected profit (solution value) and computing time. In Section 2.7.2, we report *relative gaps* measuring the relative differences between exact and heuristic solution values (for the instances we focus on, this corresponds to optimality gaps). On the contrary, *gaps* reported in Section 2.7.3 correspond to the usual gaps given by general-purpose solvers (relative differences between lower and upper bounds) but here after solving the *reduced formulation*. Hence, a zero gap does not imply an optimal solution to the original formulation.

Infrastructure and computing time budget. We perform computations on the Linux version of ILOG CPLEX 12.10 running on an Intel Core i7-7800X CPU at 3.50 GHz. Given the strategic nature of our problem, we consider a 10-hour computing time budget reasonable. We impose this time limit but note that all instances in Sections 2.7.2, 2.7.3.1 and 2.7.3.2 are solved in less than two hours.

2.7.2. Comparison of Exact and Heuristic Solution Approaches

In this section, we compare the quality of solutions obtained with our heuristic with optimal solutions computed with a standard Benders decomposition method (Section 2.5.1). We refer to the exact method as BA, and to the two variants of our heuristic as H-A (analytical) and H-S (scenario-based). We use the setup described in Section 2.7.1 which results in 66 different model configurations. We pseudo-randomly generate ten examples for each configuration.

The required number of scenarios for BA is assessed in a stability analysis (Kaut and Wallace, 2003) using a computing time limit of two hours per instance. We use the instances for which we obtain stable results with 5,000 scenarios (610 instances out of a total of 660). Moreover, we focus on those that lead to non-trivial solutions

(i.e., non-zero profits which corresponds to 350 instances). To allow for a fair comparison, we compute ρ parameters for H-S using the same set of scenarios as BA. We report detailed results for $|S| = 5000$ in Table A.6 of Appendix A.5.

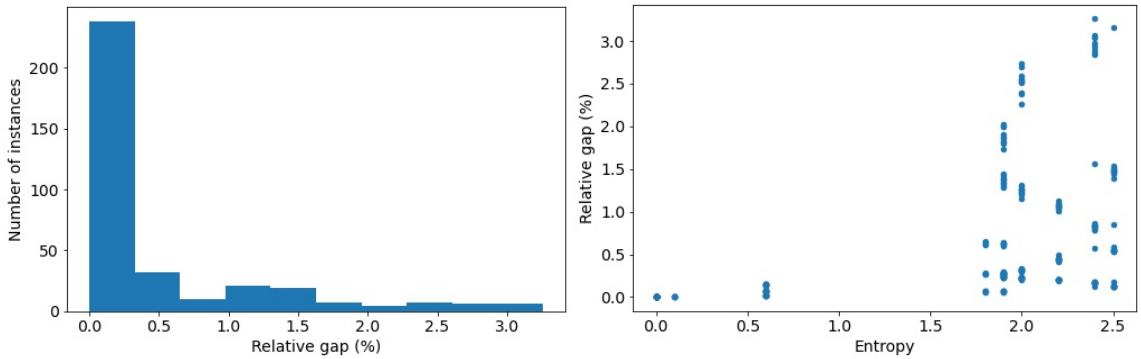
In Figure 2.2a, we show a histogram of the relative gap between objective values obtained by BA and H-S. More than 200 instances (57% of total) show a relative gap lower than 0.5% and 87 (24% of total) are solved to optimality. To shed further light on these results, Figure 2.2b displays a scatter plot relating relative gap to the entropy of each instance. We compute the latter as in Legault and Frejinger (2023), defined as $\sum_{a \in A} \mathrm{v}_a \log(\mathrm{v}_a)$, where A identifies the set of possible agreements (shippers and customer with the corresponding service levels and price assignments), and v_a, the associated probabilities. The results clearly show that H-S achieves high-quality solutions for all instances but the variance in relative gap is larger for instances with higher entropy. As the heuristic is based on a reduced model that does not explicitly take into account the scenarios, it is expected that the approximation is poorer for higher entropy (in this case more scenarios are required for a high-quality SAA). We note that the demand model approaches a uniform distribution for very high entropy values (very large β values), whereas this uninformative distribution is not of interest for our approach. Moreover, we note that the exact method does not reach stable results for such cases, even for 5,000 scenarios (see results in Appendix A.5).

In terms of computing time we note that *H-S* solves each instance in less than one second, a time that is negligible in comparison to the exact method (see Table A.7 in Appendix A.5 for details).

Finally, our computational results show negligible differences in both solution quality and computing time between *H-S* and *H-A* (see detailed results in Tables A.7 and A.8 in Appendix A.5). We use *H-A* in the subsequent experiments as the ρ parameters are computed analytically and hence are more accurate.

2.7.3. Impact of Model Parameters on Performance

This section analyzes the impact of key parameters on optimal expected profit and computation time, and assesses the maximal size of instances that can be solved to optimality. We proceed in three steps. First, we analyze how profit varies in response



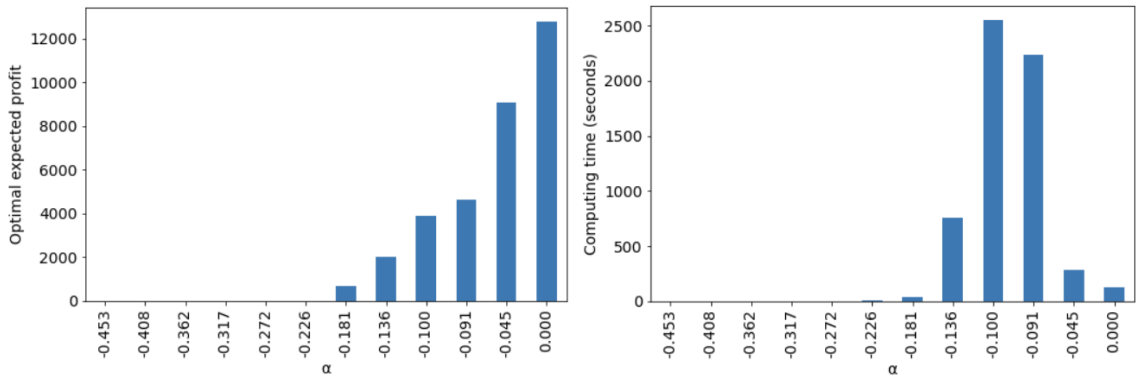
(a) Histogram of relative gap **(b)** Scatter plot: Entropy vs relative gap

Fig. 2.2. Results on relative gap and entropy

to changes in the RUM parameters (Section 2.7.3.1). For the sake of interpretability, we base this analysis on one example instance. Second, we capitalize on the CFLP structure of our problem and generate multiple instances with varying characteristics (Section 2.7.3.2). Third, we assess how computation time scales with problem size (Section 2.7.3.3). More precisely, we solve instances from three broad classes defined based on the demand model and CFLP parameters, and where instances differ in size inside these classes.

2.7.3.1. Interpreting the Impact of the Demand Model Parameters. In this section, we analyze the impact of (i) the price sensitivity parameter value α on the optimal expected profit, and (ii) the level of uncertainty captured by the scale parameter β. We also report the computing times. For ease of interpretation, we focus on a single instance. Hence, the results are for illustrative purposes. Appendix A.3 reports details about the choice of parameter values.

In Figure 2.3a, we show the optimal expected profit for α values ranging from high price sensitivity of customers ($\alpha = -0.453$) to price insensitivity ($\alpha = 0$). Note that β is fixed to 1. As expected, profit increases with decreasing price sensitivity. With highly price sensitive customers (in these results $\alpha \leq -0.226$), the optimal expected profit is zero. For this case, LP does not tender any service offer as the expected costs exceed the expected revenue. These illustrative results highlight the importance



(a) Optimal expected profit **(b)** Computing time

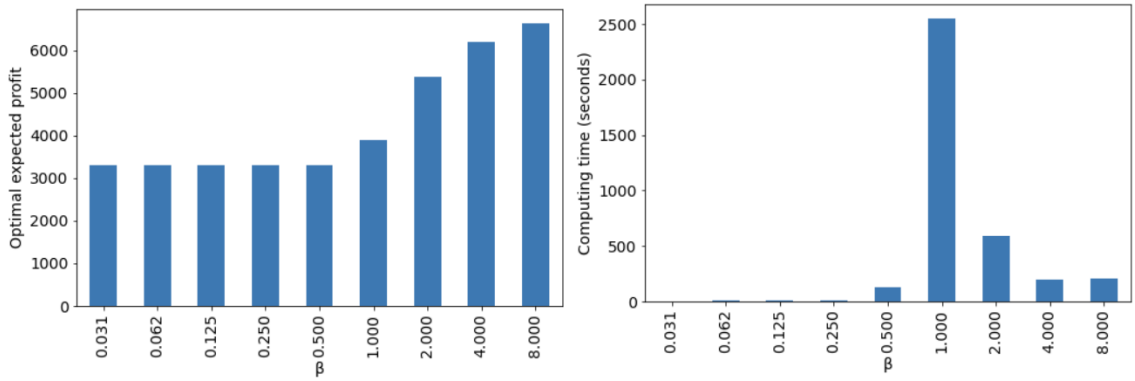
Fig. 2.3. Optimal expected profit and computing time according to α

of modeling price sensitivity as it can have a major impact on the expected profit. Underestimating its value may lead to overestimating the expected profit, and in turn, to strategic facility location decisions that do not lead to profitable operations.

In Figure 2.3b, we display a similar plot for the computing time. We observe that a peak occurs for α close to -0.1. In this case, the problem becomes harder to solve due to a symmetry issue that arises when $\rho_{nk}^{mp} q^p d_j$ in the objective function is similar for all price levels p. The analysis of the computing time for larger instances in Section 2.7.3.3 focuses on instances displaying such symmetry issues.

We now turn our attention to the impact of uncertainty. For this purpose, we vary the scale parameter of the Extreme Value type I distributed random terms. In Figure 2.4a, we report the optimal expected profit for different values, ranging from $\beta = 0.031$, representing a close to degenerate distribution, to $\beta = 8$. Note that we keep α fixed to -0.1. As uncertainty increases, we see a similar effect as that of decreasing price sensitivity, namely an increase in optimal expected profit. At the extreme, for large enough values of β, the distribution is close to uniform. It is then equally likely to choose the opt-out option (i.e., as if it does not depend on price). In other words, the deterministic part of the utility does not have an impact. We conclude that there is a trade-off between the magnitude of the deterministic part of the utility and the scale of the random term. We illustrate this in Figure 2.5 where

we show how the optimal expected profit varies for different values of α and two alternatives about β. When $\beta = 32$, there is very little variation in optimal expected profit for all values of α (we note a slight decline as the magnitude of α increases). On the contrary, when $\beta = 1$, the price sensitivity impacts the expected profit, albeit less severely than in the degenerate (deterministic) case.



(a) Optimal expected profit **(b)** Computing time

Fig. 2.4. Optimal expected profit and computing time according to β

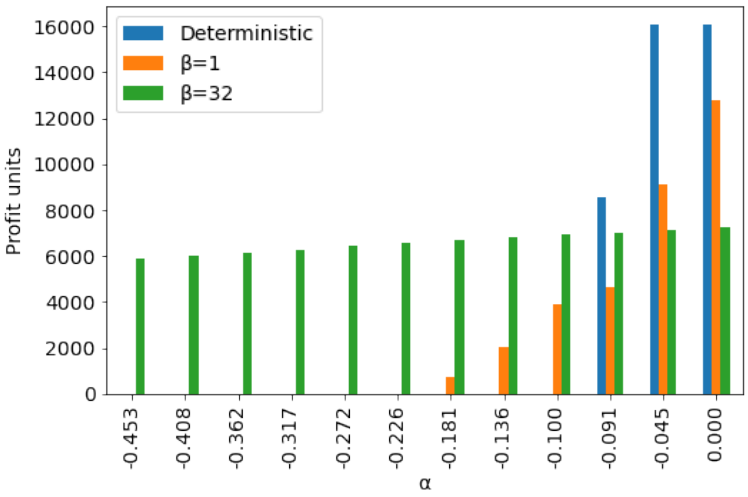


Fig. 2.5. Optimal profit according to α for $\beta = 1$, $\beta = 32$ and a deterministic utility

Finally, in Figure 2.4b, we show how computing time varies as a function of β values. We note a peak for $\beta = 1$. For the same reason that we explained in the analysis of α values, this is due to symmetry issues arising around $\beta = 1$.

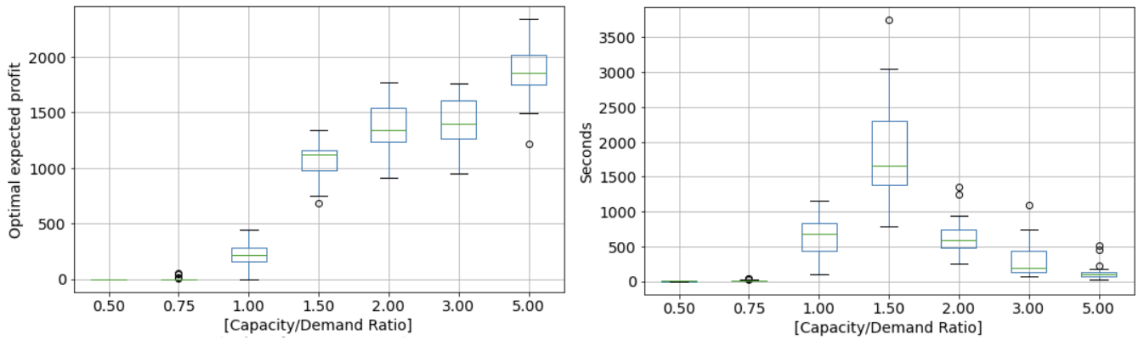
The preceding illustrative results highlight that the optimal expected profit can display important variations depending on shippers' price sensitivity and on uncertainty in utilities. It is therefore important to take such considerations into account when making strategic location decisions. The formulation we propose can model settings ranging from full information (deterministic) to no information (uniform distribution). These settings also impinge on computing times and tractability. The following sections focus on computational aspects, starting with the impact of the CFLP parameters.

2.7.3.2. Impact of CFLP Parameters. In this section, we vary the ratio parameter defined by Klose (2018) relating the total capacity of the facilities and the total demand of the customers. Whereas it was fixed to 2 in the previous section, we now let it vary from a highly constraining capacity (ratio equals 0.5, meaning that total capacity is 50% of total demand) to a high excess capacity (ratio equals 5). We note that Klose and Drexl (2005) and Cornuejols et al. (1991) also used some of the values in this range in their numerical results.

In our experiments, we treat total demand as fixed and only let the capacity vary. For each value of the ratio, we pseudo-randomly generate 20 examples.

We report the results in Figures 2.6a and 2.6b. We see that the average optimal expected profit increases with the ratio. When the ratio increases, fewer facilities are required, which affects the profits given the high impact of fixed costs. However, we should expect a decrease in profits beyond some value of the ratio. In this case, a single (the cheapest) facility captures all demand and subsequent increases in the ratio value only make the cost of this single facility increase.

In Figure 2.6b, we observe that the model is harder to solve and shows more variable computing times when the total potential capacity is 1.5 times the total demand. Instances with very loose or very constrained capacities are easily solved. We note that instances with ratio less than one are frequently the easiest to solve by leading to trivial solutions where the LP earns zero expected profit.



(a) Optimal expected profit **(b)** Computing time (seconds)

Fig. 2.6. Optimal expected profit and computing time according to ratio between total capacity and total demand R

Our problem features similarities with CFLP and, for certain parameter settings, we observe patterns in computing times that are similar to those reported about CFLP. However, the profit maximization objective and the inner structure of our problem make it quite distinct from CFLP. For instance, capacity / demand ratios less than one are not considered in CFLP since there is typically a demand satisfaction constraint. We also note that certain settings lead to trivial solutions requiring negligible computing times. Consider for example instances where I and J have similar cardinalities. If a tight ratio between total capacity and total demand holds, we always obtain a trivial solution in our problem, while this is not the case for results reported about CFLP.

2.7.3.3. Impact of Problem Size. To analyze the impact of problem size on a diverse set of instances, we define three classes of instances based on the demand model and CFLP parameters. Inside these classes, instances differ essentially with respect to the number of customers, facilities, and price levels.

The first two classes comprise larger instances than in the previous sections. Those are generated in the same way through pseudo-random sampling, using the generator of Klose (2018). The first class, said *hard*, is created with parameter values that are leading to large computing times in previous sections, that is $\alpha = -0.1$, $\beta = 1$, and $R = 1$. The second class is said *easy*. Although its instances are harder

than those in the previous sections, they are created with parameter values leading to computing times that are lower in comparison with those in the hard class, that is $\alpha = -0.15$, $\beta = 0.5$, and $R \in \{5, 10\}$. Moreover, for the latter instances we do not impose a minimum demand level (i.e., $l_n^{mp} = 0$), hence relaxing constraints (2.16e). We pseudo-randomly generate samples of size 20 for each instance in the easy and hard classes.

The third class, *ORLib*, contains instances that we create based on *OR-Library* (Beasley, 1990). Specifically, we select problem set A and problems IV-XIII (we do not include the smaller instances I-III). Since the instances in set A are very large, we create an additional set named A-reduced containing instances identical to those in A, except for their reduced size. In all of the instances, each service level is associated with a different set of price levels (see Table A.5 in Appendix A.3 for details) and there is only one customer category (i.e., $|K| = 1$). Each instance has parameters d_j, f_i, u_i, and c_{ij} already defined (see Appendix A.3) for the specific values).

We use $\alpha = -0.1$ and $\beta = 2$, while the rest of parameters in the RUM model remain as in previous sections. Moreover, we use $l_n^{mp} = 0$ as in the easy class of instances above.

Table 2.2 reports statistics about the performance of our formulation for each set of instances in the hard and easy classes. We show the minimum, average, and maximum computing times (in seconds) over the 20 generated examples in columns Min., Avg. and Max., respectively. The number of cases out of 20 reaching a zero gap within the allocated time is reported in column #Solved. As expected, computing time increases with the size of the problems measured by the number of customers and facilities. It does so linearly. On the contrary, computing time increases exponentially with the number of price levels. This is very likely caused by the escalation in the symmetry issues identified in Section 2.7.3.1 as the number of price levels increases. We cannot solve three out of 20 generated examples of the largest hard instance within the time limit. For those three instances, the average gap is 11.2%.

Noteworthy is the significant difference in computing times reported in Table 2.2 between the hard and easy classes. For example, the largest instance in the easy class (i.e., $|I| = 8, |J| = 200, |P| = 10, R = 5$) is solved 3.5 times faster, on average,

$\|I\|$	$\|J\|$	$\|P\|$	α	β	R	Computing Time (sec.) Min.	Avg.	Max.	# Solved
Hard Setting									
4	80	3	-0.1	1	1	101.0	201.6	402.0	20
4	80	4	-0.1	1	1	941.0	1528.8	2340.0	20
4	80	5	-0.1	1	1	3170.0	5103.0	7970.0	20
5	80	5	-0.1	1	1	2470.0	5489.0	8690.0	20
6	80	5	-0.1	1	1	3080.0	6421.5	11500.0	20
5	100	5	-0.1	1	1	6460.0	9371.0	11900.0	20
5	120	5	-0.1	1	1	6550.0	9911.0	12400.0	20
4	140	5	-0.1	1	1	5580.0	9779.5	14700.0	20
5	140	5	-0.1	1	1	9080.0	13504.0	19000.0	20
6	140	5	-0.1	1	1	14000.0	22615.0	34700.0	20
7	140	5	-0.1	1	1	15200.0	26452.9	35500.0	17
Easy Setting									
6	180	5	-0.15	0.5	5	386.0	751.8	1870.0	20
6	200	5	-0.15	0.5	5	239.0	375.0	535.0	20
6	200	10	-0.15	0.5	5	554.0	1137.1	1770.0	20
8	200	5	-0.15	0.5	5	132.0	1615.9	4790.0	20
8	200	10	-0.15	0.5	5	2100.0	7428.0	32900.0	20
10	300	5	-0.15	0.5	5	2560.0	8969.5	17900.0	20
6	180	5	-0.15	0.5	10	77.8	199.3	285.0	20
6	200	5	-0.15	0.5	10	202.0	333.7	554.0	20
6	200	10	-0.15	0.5	10	700.0	1284.0	3110.0	20
8	200	5	-0.15	0.5	10	556.0	940.0	1700.0	20
8	200	10	-0.15	0.5	10	1360.0	3422.5	8200.0	20
10	300	5	-0.15	0.5	10	4440.0	7923.2	13800.0	20

Table 2.2. Computing times for hard ($\alpha = -0.1, \beta = 1, R = 1$) and easy ($\alpha = -0.1, \beta = 0.5, R \in \{5,10\}$) classes of instances

than the largest in the hard class (i.e., $|I| = 7, |J| = 140, |P| = 5$). We also note that changing the value of parameter R from 5 to 10 makes the problem easier to solve, as also seen in Section 2.7.3.2.

Next, we turn our attention to results for the *ORLib* class of instances. Table 2.3 reports the performance related to problems IV - XIII, and columns $|P| = 5$,$|P| = 10$, and $|P| = 15$ show the computing times for the corresponding size of the price level set. We do not report gaps as they are zero for all instances. As before, we observe that computing time increases faster with the number of price levels than the number of facilities.

Table 2.4 reports computing times and gaps for the hardest instances, namely those of problem set A. We note that the original instances are too hard, as shown by the large gaps at timeout. However, the relatively large instances in class A-reduced

can be solved, even for 15 price levels. We note that the computing time varies with the capacity of the facilities which is consistent with the other findings (computing time varies with parameter R).

We conclude from this section that the performance of our heuristic depends on parameter settings, chiefly those pertaining to CFLP and demand model. In view of the intended strategic use of the model and the high-level representation of the pricing problem, the number of price levels is bound to remain limited and we do not expect this situation to cause prohibitive difficulties in practice.

2.8. Case Study

This section presents a realistic case study based on the parcel delivery operations of our industrial partner – Purolator – in Montreal. We study the case when the LP, Purolator, seeks to conclude agreements with two new shippers. Concluding such agreements would require expanding the capacity of the existing facilities and our model is used to assess profitability.

To ensure the confidentiality of sensitive information, we transform the problem while preserving realism of the results. We do not distinguish shipper and customer categories, and we assume that there is a single customer category and that demand in each zone is uniformly distributed over the two shippers. Price levels offered to final customers differ from actual values.

We assume that a logit model holds and that its utilities, $U_n^m = \sum_{p \in P_n^m} \alpha_n^m q_n^{mp}$, only depend on the price associated with the service level. Also, we assume that the utility of the opt-out alternative is $U_n^0 = 0$. This is not a strong assumption as only differences in utilities matter.

We consider a set of 71 demand zones and a possible capacity expansion of two terminals – A and B. The location of the demand zones and the two terminals are illustrated in Figure 2.7. The size of the demand points (depicted with circles) is proportional to the amount of demand in the corresponding zone. For both terminals, the LP can add capacity of five million parcels per year.

					Computing Time (sec.)		
Problem Set	$\|I\|$	$\|J\|$	u_i	f_i	$\|P\| = 5$	$\|P\| = 10$	$\|P\| = 15$
IV	16	50	5000	7500	5.6	12.1	25.9
IV	16	50	5000	12500	5.3	11.9	24.6
IV	16	50	5000	17500	5.1	12.5	18.3
IV	16	50	5000	25000	5.0	13.0	22.6
VI	16	50	15000	7500	4.5	9.9	22.5
VI	16	50	15000	12500	4.4	10.3	18.8
VI	16	50	15000	17500	5.3	10.4	23.3
VI	16	50	15000	25000	6.3	11.0	22.7
VII	16	50	58268	7500	6.1	16.5	35.6
VII	16	50	58268	12500	6.1	15.2	30.2
VII	16	50	58268	17500	6.6	18.1	33.6
VII	16	50	58268	25000	5.6	13.2	30.5
VIII	25	50	5000	7500	12.9	31.8	54.8
VIII	25	50	5000	12500	13.9	34.7	55.4
VIII	25	50	5000	17500	13.0	32.5	50.1
VIII	25	50	5000	25000	13.6	28.7	58.7
IX	25	50	15000	7500	13.3	21.2	49.6
IX	25	50	15000	12500	10.7	25.4	50.8
IX	25	50	15000	17500	19.4	26.6	46.7
IX	25	50	15000	25000	21.4	27.3	40.5
X	25	50	58268	7500	13.2	44.1	91.3
X	25	50	58268	12500	11.7	41.0	107.0
X	25	50	58268	17500	21.8	51.9	141.0
X	25	50	58268	25000	14.5	46.4	101.0
XI	50	50	5000	7500	31.0	72.5	164.0
XI	50	50	5000	12500	36.2	206.0	157.0
XI	50	50	5000	17500	40.2	82.8	160.0
XI	50	50	5000	25000	39.7	74.3	131.0
XII	50	50	15000	7500	29.7	83.5	162.0
XII	50	50	15000	12500	39.1	98.3	149.0
XII	50	50	15000	17500	58.7	72.0	186.0
XII	50	50	15000	25000	57.8	68.6	145.0
XIII	50	50	58268	7500	86.8	434.0	729.0
XIII	50	50	58268	12500	87.5	371.0	687.0
XIII	50	50	58268	17500	91.7	371.0	1060.0
XIII	50	50	58268	25000	93.1	287.0	530.0

Table 2.3. Computing times for *ORLib* class instances adapted from problems *IV* to *XIII* in Beasley (1988)

The total demand of the two shippers represents 60% of this total additional capacity. However, the fixed cost of implementing such an expansion in terminal B is 3.3 times higher compared to terminal A.

As reported in Table 2.5, we consider four service levels and seven price levels each. This results in 28 alternatives. The service levels are defined based on lead

Problem Set	$\|I\|$	$\|J\|$	u_i	$\|P\|$	Time (sec.)	Gap
A	100	1000	8000	5	36000	27%
A	100	1000	10000	5	36000	80%
A	100	1000	12000	5	36000	69%
A	100	1000	14000	5	36000	73%
A-reduced	25	500	8000	5	2370	0%
A-reduced	25	500	10000	5	1890	0%
A-reduced	25	500	12000	5	2400	0%
A-reduced	25	500	14000	5	2560	0%
A-reduced	25	500	8000	10	12000	0%
A-reduced	25	500	10000	10	9800	0%
A-reduced	25	500	12000	10	11400	0%
A-reduced	25	500	14000	10	9610	0%
A-reduced	25	500	8000	15	23600	0%
A-reduced	25	500	10000	15	18500	0%
A-reduced	25	500	12000	15	24300	0%
A-reduced	25	500	14000	15	22900	0%

Table 2.4. Computing times and gaps for large instances in *ORLib* class, adapted from problem sets A and A-reduced

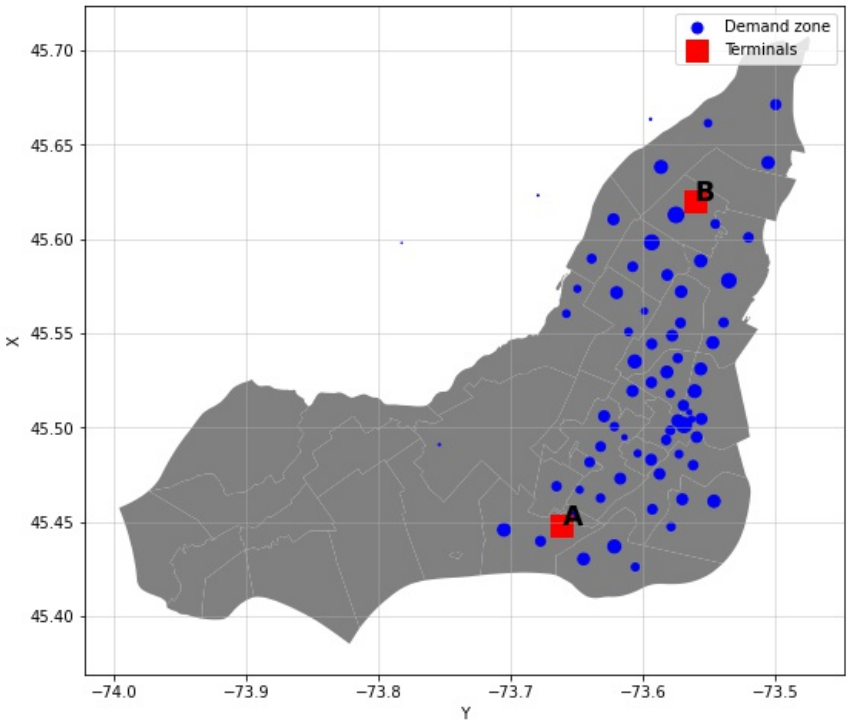


Fig. 2.7. Demand zones and terminals in the case study

times. Lower lead times yield a higher impact on operating costs as reflected by the Cost Scale Factor parameter. In the table, we report one price sensitivity parameter – $\alpha_0^{mp}, \alpha_1^{mp}$ – for each shipper. All price and service level alternatives are available regardless of the demand level of the shippers, thus $l_n^{mp} = 0$.

Service Level	Set of Price Levels	Cost Scale Factor	$\{\alpha_0^{mp}, \alpha_1^{mp}\}$
Next two days - before 21pm	$\{18, 24, 33, 37, 41, 51, 52\}$	1.0	$\{-0.11, -0.13\}$
Next day - before 21pm	$\{19, 24, 33, 37, 42, 52, 55\}$	1.25	$\{-0.0825, -0.13\}$
Next day - before 12 pm	$\{21, 27, 37, 42, 47, 59, 65\}$	1.5	$\{-0.055, -0.13\}$
Next day - before 9:00 am	$\{52, 56, 68, 73, 79, 91, 97\}$	1.6	$\{-0.055, -0.13\}$

Table 2.5. Price and service level data in the case study

In this setting, the LP concludes agreements with both shippers. To satisfy the agreements, the LP adds a capacity of five million parcels to each terminal. The LP and Shipper 0 agree on *Next day - before 12 pm* deliveries at \$47 per parcel, whereas with Shipper 1, the agreement is on *Next two days - before 21pm* at \$24 per parcel.

The solution and the expected optimal profit depend on the value of the price sensitivity parameters. For example, if Shipper 1 becomes 10% more price sensitive, the expected optimal profit decreases by approximately 3%. If Shipper 1 is 3.4 times more price sensitive, it is not profitable for the LP to conclude an agreement with that shipper, as the additional costs surpass the revenues. In that case, the LP would only require to expand the capacity of Terminal A to serve the additional demand coming from Shipper 0. Terminal B remains with the same capacity level and is not used to service any of the shippers in this problem.

Through the agreements, the optimal profit also depends on the capacity of the facilities. On the one hand, consider the original example, if the LP can only add 50% of the planned expansion (with a proportional reduction in the fixed cost), the LP prioritizes an agreement with Shipper 0, with the same deal as before. In this case, there is no agreement with Shipper 1, as there is insufficient capacity. Therefore, the optimal expected profit reduces by 27% compared to the previous value. On the other hand, consider the example with Shipper 1 being more price sensitive. Then, if Terminal A had enough capacity to serve demand from both shippers (6 million parcels per year), then an agreement with Shipper 1 is still possible. However, due to

the higher price sensitivity of Shipper 1, an agreement is reached at the same service level but at the lowest price.

In summary, increased price sensitivity decreases expected profits, but not proportionally. Moreover, capacity impacts agreements and hence expected profitability. Finally, we note that it takes less than one second to solve the instances related to this case study.

2.9. Conclusion and Perspectives

We addressed a strategic decision-making problem for an LP seeking facility location decisions that lead to profitable operations. In this context, we model at a high level the decisions related to price and service levels, in addition to facility locations. In this strategic setting, the LP has imperfect knowledge of the preferences of the shippers and their customers that impact its profit.

Offering a first attempt to address this problem that, to the best of our knowledge, has not been studied before, we proposed a reduced single-level MILP formulation for a bilevel stochastic formulation. We grounded this formulation in a simulation-based model, and three key results arose from the analysis of its structure: First, we found that the problem of the follower can be formulated as a continuous, unconstrained optimization model. Second, we defined conditions to discover trivial solutions. Third, we reduced the size of the problem by including a parameter that approximately condenses the information of all scenarios. We can easily compute this parameter outside of the model in a preprocessing phase. This resulted in a model that can be solved by a general purpose solver for instances of sizes that are relevant in practice.

In addition to optimizing facility location decisions, our methodology allows the LP to anticipate the impact of its decisions about price and quality of service. For instance, solutions can prepare an LP for a transportation procurement process by providing tailored prices and service levels depending on the shippers' profiles. Our extensive experimental results showed that some facility location decisions lead to zero expected profit for the LP, as the resulting costs would exceed revenues.

We also observed that the expected profit is highly dependent on the demand model parameters. Finally, achieved computing times were relatively large for some parameter configurations due to symmetry issues. Such issues worsen when the number of price levels increases.

We envisage pursuing research along three directions. The *first* direction concerns extensions to our model. Instead of setting a service level and a price for each shipper independently, the LP's policy might address the aggregated demand of several shippers. Also, service levels could be more finely defined. Thus, given a set of attributes and the preferences of the customers towards them, a decision would be made on the optimal level of each attribute in a service level offered to a shipper. The *second* direction pertains to the estimation of the demand model. In view of the high dependency of the solutions on the latter (in our case a RUM model), parameter estimation might gainfully account for partial availability of data, as important attributes may be latent. This would also make it possible to test the methodology on real data. *Finally*, additional work should be dedicated to improving the algorithmic performance. Some parameter configurations lead to symmetry issues, and a column generation approach would seem appropriate in these circumstances as well as enhancements to the model to include symmetry-breaking inequalities. Another potentially promising avenue is the use of an appropriate metaheuristic to solve instances of very large size. Camacho-Vallejo et al. (2024) provide a comprehensive review of metaheuristics for bilevel optimization. The most popular approach consists in solving the leader problem and then evaluating the follower problem. In our case, however, we believe that it is more promising to solve the single-level reformulation as it exploits the optimal solution structure of the followers' problem. An important challenge resides in designing the local search which is not straightforward due to the complex cost structure of our problem.

Chapter 3 Second article

A Pattern-based Reformulation of a Profit-maximizing Facility Location Problem with RUM Followers

by

David Pinzon Ulloa[1], and Emma Frejinger[1]

([1]) Département d'informatique et de recherche opérationnelle, Université de Montréal, Montréal, Canada.
CIRRELT, Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation

Working paper to be submitted to a journal.

The main contributions of the authors for this article are as follows.

- David Pinzon proposed the model formulations, designed, implemented, ran the experiments, and wrote the article.
- Emma Frejinger provided advice on the methodology and contributed to the mathematical formulations and writing of the article.

ABSTRACT. We propose a reformulation of a logistics provider's (LP) profit-maximizing facility location problem with random utility-maximizing followers. This reformulation aggregates scenarios displaying computationally similar preference patterns. These patterns are defined by combinations of price and service level assignments to shippers and customers that lead to the same agreements with the logistics provider. This reformulation (BA-R) achieves good performance in instances with low levels of uncertainty. The BA-R's computing time is smaller by several orders of magnitude than that of the exact approach presented in Pinzon Ulloa et al. (2024). Moreover, thanks to the reduced computing time, we can extend the set of instances used for benchmarking the performance of the heuristics proposed in the latter paper.
Keywords: Facility location, logistics, pricing, pattern reformulation, profit maximization, random utility maximization models

3.1. Introduction

The continuous growth of online shopping, urbanization, traffic congestion, and customer requirements impose challenging conditions for logistic operations. These conditions have stimulated shippers to outsource non-core business operations, such as transportation, to specialized logistics providers (LP) (Langley and Infosys, 2019). The shippers aim to offer higher customer service levels and reduce logistics costs. LPs aim to secure their own maximal profits when they offer their services to shippers—their clients. In turn, these profits are affected by the services promised in the agreements with their clients.

This work addresses an LP's strategic decision-making problem of ensuring profit-maximizing operations through facility location and price-service level decisions. Profits depend on the revenues and costs resulting from the agreements with the shippers. These revenues depend in turn on the demand captured from the shippers

and the price and service level specified in the agreements. The costs result from the logistic decisions to fulfill those agreements. They depend on the service levels agreed upon with the shippers and on the locations of their customers. This problem is challenging due to the interdependence between the LP, the shippers, and their customers. While the agreements are concluded between the LP and the shippers, the costs are related to the characteristics of the shippers' customers. However, the LP has incomplete information about this at the strategic level.

Relations with Existing Literature. Pinzon Ulloa et al. (2024) present a choice-based optimization problem, more precisely, a two-stage stochastic bilevel facility location and pricing problem including random utility-maximizing shippers and a profit-maximizing LP. The utility of the shippers accounts for their own preferences in regard to the combinations of price and service level offered by the LP and for the preferences of their customers in regard to the service levels. Compared to other works, they allow for a complex relationship between price and costs, as well as for endogenous uncertainty around the costs resulting from customer behavior.

Pinzon Ulloa et al. (2024) propose an exact and a heuristic approach to solve the model they present. The two approaches are rooted in a sample average approximation (SAA) formulation using a sample of pseudo-randomly simulated utilities (scenarios) to approximate the required choice probabilities (see Pacheco Paneque et al., 2021, for details). The exact approach, denoted by BA, relies on Benders decomposition to solve the SAA formulation for a given set of scenarios, whereas the heuristic approach solves an approximate reduced, single-level formulation. Note that they distinguish two variants of the heuristic: H-A, which uses analytical probabilities resulting from the assumption of a logit choice model, and H-S, which uses the information from the simulated utilities. Both approaches, the exact and the heuristic, are established on the foundation of a simulation-based reformulation. The experimental results show that the heuristics can achieve optimality gaps lower than 5%. Yet, the accuracy of H-A was only evaluated on a set of small instances, due to BA requiring large computing times, especially when large sets of scenarios are needed to obtain stable optimal solutions. The main difficulty experienced with the

BA approach relates to the increases in computing time associated with increasing numbers of scenarios.

Benati and Hansen (2002), Haase (2009), and Zhang et al. (2012) present, in connection with choice-based optimization problems, linear reformulations of a nonlinear maximum capture problem. Haase and Müller (2014) concludes that Haase (2009) offers the most promising proposition. However, the latter is tied to logit models and the use of analytical choice-probabilities. Other authors propose approximations to the original nonlinear choice-based optimization problems based on samples of observations of the stochastic terms. Those are not restricted to specific choice models and use simulated utilities to approximate the choice probabilities. See, e.g., Haase and Müller (2013); Pacheco Paneque et al. (2021); Legault and Frejinger (2023); Pinzon Ulloa et al. (2024). These approaches asymptotically converge to the true solution when the size of the sample increases. An alternative approach consists in reducing the number of scenarios needed (see e.g., Chen, 2019; Bertsimas and Mundru, 2023; Emelogu et al., 2016) instead of reformulating existing models.

Among the publications addressing the choice-based facility location problems, Legault and Frejinger (2023) and the research work in Chapter 4 are our closest neighbors. Legault and Frejinger (2023) propose a strategy that identifies scenarios with identical preference profiles and aggregate them into patterns. In contrast to this paper, they do not address capacitated facility location problems and assume a submodular objective function. Our work in Chapter 4 proposes a reformulation that applies a preference profile aggregation method similar to that of Legault and Frejinger (2023) but does so to the capacitated facility location problems. However, the model formulations in Chapter 4 do not impose capacity constraints in each scenario or pattern, as we do here. Instead, they force the expected captured demands to respect capacities.

Contributions. Summing up, we seek to reformulate the simulation-based model introduced in Pinzon Ulloa et al. (2024) to obtain an exact solution approach with an improved computational performance allowing to solve larger instances. Our reformulation is inspired by the preference profile aggregation strategy introduced in Legault and Frejinger (2023). The reformulation based on aggregation of scenarios

with similar preference profiles outperforms the computing time of BA by approximately 2-3 orders of magnitude. The former makes it possible to provide additional evidence on the accuracy of H-A and H-S in comparison with exact solutions.

Content. The remainder of the paper is structured as follows: Section 3.2 introduces the reformulation based on the patterns identified from the scenarios. Section 3.3 reports on the experimental results collected on this reformulation and provides new evidence on the accuracy of the H-A and H-S approaches developed in Pinzon Ulloa et al. (2024). Finally, Section 3.4 concludes.

3.2. Pattern-based Reformulation

This section presents the reformulation aggregating scenarios to reduce the number of decision variables and constraints. We start by recalling some concepts from Pinzon Ulloa et al. (2024) and defining new ones. Then, we explain the process leading to the pattern-based reformulation.

Let us recall that in Pinzon Ulloa et al. (2024), an LP seeks profitable agreements with shippers by proposing a set of alternatives given by the combinations of price ($p \in P$) and service levels ($m \in M$). There, y_n^{mp} denotes the price level p associated with service level m offered to shipper $n \in N$, z_{nk}^m the service level m offered to customers in category $k \in K_n$ of shipper n. In addition, r_i denotes the decision whether to open or not a facility in location i within a set of facility locations I. Each facility has an associated capacity u_i and a fixed cost of use f_i. Also, x_{nks}^m identifies the decision of shipper n to accept the service level m for its customer category k in a given scenario $s \in S$. Note that this decision depends on the utility that service level m and price p provide to the customers in category k of a shipper n, for the scenario s, denoted by U_{nks}^{mp}. The utility of the *opt-out* alternative, for a given scenario s, is denoted by U_{nks}^0. Moreover, recall that each scenario $s \in S$ corresponds to a realization of the random utilities $U_{nk}^m(\mathbf{y}),\ \forall n \in N, m \in M^0$ and $k \in K_n$, given price level decisions $\mathbf{y}$. In this case, M^0 denotes the set of available service levels including the opt-out alternative.

The key step in introducing the reformulation is defining the underlying patterns. We seek patterns identifying a subset of similar feasible agreements. An agreement between the LP and the shippers is feasible if it satisfies the shippers and their customers. More precisely, an agreement with a shipper is feasible if the utility associated with the price and service level offered by the LP is superior to the utility of the opt-out. Formally, let $\delta_s \in \{0,1\}^{|T|},\ s \in S$ be a 0-1 vector of dimension $|T| = |N||K||M||P|$ representing potential service level and price assignments included in the agreements between the LP and the shippers for a scenario s (or simply, potential agreements). Recall that an assignment can result in an agreement if, for the corresponding scenario s, $U^{mp}_{nks} > U^{0}_{nks}$. Hence, each element $\delta_{st} = \mathbb{1}(U^{mp}_{nks} > U^{0}_{nks})$, where the position $t \in \{1,...,|T|\}$ is computed as follows: $t = n \times (|K| \times |M| \times |P|) + k \times (|M| \times |P|) + m \times |P| + p$, for given indices n, k, m and p. Table 3.1 shows the profile of potential agreements for four scenarios, in an instance with $N = \{0,1\}$, $K = \{0,1\}$, $M = \{0\}$, and $P = \{0,1\}$ (hence, $|T| = 8$). For example, δ_0 shows that in scenario $s = 0$, it is only possible to reach an agreement with shipper $n = 0$ where the shipper accepts a service level $m = 0$ for its customers in category $k = 0$ with a price level $p = 0$ (hence $\delta_{00} = 1$, $\delta_{0t} = 0$ for $t \neq 0$). Note that a potential agreement profile in a scenario s includes all the combinations of price service level for which $U^{mp}_{nks} > U^{0}_{nks}$ for a shipper n and customers in category k. In other words, all price and service level combinations for which shipper n includes customers in category k in an agreement with the LP. As an example, in scenario $s = 2$ of Table 3.1, both price levels $p = \{0,1\}$ are candidates for shipper $n = 0$ and customers in category 0 ($\delta_{st} = 1, t = \{0,1\}$).

	δ_s							
s	0	1	2	3	4	5	6	7
0	1	0	0	0	0	0	0	0
1	0	0	1	0	1	0	0	0
2	1	1	0	0	1	0	1	0
3	1	0	0	0	0	0	0	0

Table 3.1. Example of potential agreements

The BA approach in Pinzon Ulloa et al. (2024) solves $|S|$ subproblems each involving a capacitated facility location problem. However, inspired by Legault and

Frejinger (2023), we note that we can group scenarios $s \in S$ sharing a common set of potential agreements. More precisely, two scenarios $s_1, s_2 \in S$ are included in the same pattern a (or agreement pattern) if and only if $\delta_{s_1} = \delta_{s_2}$. Thus, according to the potential agreements given in Table 3.1, $s = 0$ and $s = 3$ belong to the same pattern. Crucially, a subset of scenarios assigned to the same pattern will result in the same optimal solution for decision variables x^m_{nks}. Let A denote the resulting partition of S containing the patterns a. A positive scalar $0 < \hat{\mathrm{v}}_a < 1$ is associated with each pattern $a \in A$ and represents the fraction of scenarios $s \in S$ sharing the same agreement pattern a and a vector $b_a \in \{0,1\}^{|T|}$. This vector b_a is defined analogously to δ_s, and encodes the common agreement pattern. That is, $b_{at} = 1$ if the utility of pattern a in the resulting position t (as defined previously) is greater than the opt-out alternative. In this case, we denote by x^m_{nka} the decision to accept, or not, the service level m assigned to customers k of shipper n in pattern a.

Using patterns we rewrite the SAA bilevel formulation (4.6) in Pinzon Ulloa et al. (2024) (note that we follow the notation introduced in Chapter 2) as an adaptation of the maximum covering problem:

$$\max \sum_{a \in A} \left(\sum_{n \in N} \sum_{k \in K_n} \sum_{m \in M_{nk}} \sum_{p \in P^m_n} \hat{\mathrm{v}}_a d_k q^{mp}_n x^m_{nka} y^{mp}_n - \sum_{m \in M} \sum_{i \in I} \sum_{j \in J} \hat{\mathrm{v}}_a c^m_{ij} d_j w^m_{ija} \right) - \sum_{i \in I} f_i r_i \tag{3.1a}$$

$$\text{s.t.} \sum_{p \in P^m_n} y^{mp}_n \leq 1, \qquad n \in N, m \in M_n, \tag{3.1b}$$

$$z^m_{nk} \leq \sum_{p \in P^m_n} y^{mp}_n, \qquad n \in N, k \in K_n, m \in M_{nk}, \tag{3.1c}$$

$$\sum_{m \in M_{nk}} z^m_{nk} \leq 1, \qquad n \in N, k \in K_n, \tag{3.1d}$$

$$\sum_{m \in M_n} \sum_{p \in P^m_n} y^{mp}_n \leq |K_n|, \qquad n \in N, \tag{3.1e}$$

$$\sum_{j} \sum_{m \in M_j} d_j w_{ija}^m \gamma^m \leq u_i r_i, \qquad a \in A, i \in I, \tag{3.1f}$$

$$x_{nka}^m \geq \sum_{p \in P_n^m} b_{nkmp}^a \pi_{nk}^{mp}, \qquad a \in A, n \in N, k \in K_n, m \in M_{nk}, \tag{3.1g}$$

$$x_{nka}^m \leq z_{nk}^m, \qquad a \in A, n \in N, k \in K_n, m \in M_{nk}, \tag{3.1h}$$

$$\sum_{i \in I} w_{ija}^m = x_{nka}^m, \qquad a \in A, n \in N, k \in K_n, j \in J_k, m \in M_{nk}, \tag{3.1i}$$

$$\sum_{m \in M_j} w_{ija}^m \leq r_i, \qquad a \in A, j \in J, i \in I, \tag{3.1j}$$

$$\sum_{k \in K_n} d_k x_{nka}^m \geq \sum_{p \in P_n^m} l_n^{mp} y_n^{mp}, \qquad a \in A, n \in N, m \in M_n, \tag{3.1k}$$

$$\pi_{nk}^{mp} \leq z_{nk}^m, \qquad n \in N, k \in K_n, m \in M_{nk}, p \in P_n^m, \tag{3.1l}$$

$$\pi_{nk}^{mp} \leq y_n^{mp}, \qquad n \in N, k \in K_n, m \in M_{nk}, p \in P_n^m, \tag{3.1m}$$

$$\pi_{nk}^{mp} \geq z_{nk}^m + y_n^{mp} - 1, \qquad n \in N, k \in K_n, m \in M_{nk}, p \in P_n^m, \tag{3.1n}$$

$$y_n^{mp} \in \{0,1\}, \qquad n \in N, m \in M_n, p \in P_n^m, \tag{3.1o}$$

$$z_{nk}^m \in \{0,1\}, \qquad n \in N, k \in K_n, m \in M_{nk}, \tag{3.1p}$$

$$x_{nka}^m \in \{0,1\}, \qquad a \in A, n \in N, k \in K_n, m \in M_{nk}, \tag{3.1q}$$

$$\pi_{nk}^{mp} \geq 0, \qquad n \in N, k \in K_n, m \in M_{nk}, p \in P_n^m. \tag{3.1r}$$

Objective (3.1a) maximizes the total expected profits. It encompasses the total revenues resulting from the accepted price and the demand captured in the agreements, the total costs from attending to the demand with the agreed-upon service level, and the fixed cost of opening the facilities to support the required operations. For complementary details on the formulation, see Section 2.5.2 in Chapter 2. The constraints serve the following purposes:

- Constraints (3.1b) ensure that at most one price level is assigned to a shipper and service level.
- Constraints (3.1c) ensure that only the service level associated with a price level can be offered to the shippers.
- Constraints (3.1d) ensure that customers are assigned no more than one service level.

- Constraints (3.1e) ensure that the number of price and service levels alternatives offered to a shipper cannot exceed the number of its customer categories.
- Constraints (3.1f) ensure that the assignment of customers respects the capacity of the facilities.
- Constraints (3.1g) link the patterns associated with combinations of price and service level with the customers. Intuitively, if customers in category k of shipper n are assigned to service level m in pattern a, then there must be a price p associated with service level m to the corresponding shipper and customer category such that $U^{mp}_{nka} > U^{0}_{nka}$.
- Constraints (3.1h) ensure that shippers cannot choose a price and service level not offered in the agreements.
- Constraints (3.1i) ensure that customers cannot be assigned to any facility if they were not included in the agreements.
- Constraints (3.1j) ensure that customers are only assigned to open facilities.
- Constraints (3.1k) ensure that the demand associated with a combination of price and service level satisfies the minimum demand requirements.
- Constraints (3.1l), (3.1m) and (3.1n) allow to linearize $z^{m}_{nk} y^{mp}_{n}$.
- Constraints (3.1o), (3.1p), (3.1q) and (3.1r) define the domain of each set of variables.

We proceed to show that the revenues (i.e., first term) in Objective (3.1a) can be formulated similarly to those in the reduced and single-level MILPs on which H-A and H-S are grounded (see Section 2.5.2 in Chapter 2). Accordingly, we write the revenues as

$$\sum_{n\in N}\sum_{k\in K_n}\sum_{m\in M_{nk}}\sum_{p\in P^{m}_{n}} \frac{|S^{mp}_{nk}|}{|S|} d_k q^{mp}_{n} \pi^{mp}_{nk}. \tag{3.2}$$

Let us recall that S^{mp}_{nk} denotes the subset of scenarios S, in which $U^{mp}_{nks} > U^{0}_{nks}$, for each n, k, m and p. More concisely, $S^{mp}_{nk} = \{s \in S : \delta_{st} = 1\}$, and let $(S^{mp}_{nk})^c$ be the corresponding complement. Hence,

$$\sum_{n}\sum_{k}\sum_{m}\sum_{p}\frac{|S_{nk}^{mp}|}{|S|}d_k q_n^{mp}\pi_{nk}^{mp} = \sum_{n}\sum_{k}\sum_{m}\sum_{p}\sum_{s\in S}\frac{1}{|S|}\mathbb{1}\big(\delta_{st}=1\big)d_k q_n^{mp}\pi_{nk}^{mp} \tag{3.3}$$

$$= \sum_{n}\sum_{k}\sum_{m}\sum_{p}\sum_{a\in A}\frac{1}{|S|}\mathbb{1}\big(b_{nkmp}^{a}=1\big)\hat{\mathrm{v}}_a|S|d_k q_n^{mp}\pi_{nk}^{mp} \tag{3.4}$$

$$= \sum_{n}\sum_{k}\sum_{m}\sum_{p}\sum_{a\in A}\hat{\mathrm{v}}_a d_k q_n^{mp}x_{nka}^{m}y_n^{mp}. \tag{3.5}$$

Equation (3.3) uses the definition of S_{nk}^{mp}. Equation (3.4) aggregates the number of cases where $\mathbb{1}\big(\delta_{st}=1\big)$ in each pattern a where $\delta_{st}=b_{nkmp}^{a}$. This leads to $\hat{\mathrm{v}}_a|S|$ scenarios. Equation (3.5) results from the equivalence between $\mathbb{1}\big(b_{nkmp}^{a}=1\big)\pi_{nk}^{mp}$ and $x_{nkm}^{a}y_n^{mp}$ as expressed in Constraints (3.1g). To show this, we distinguish two cases: (i) $b_{nkmp}^{a}=0$ and (ii) $b_{nkmp}^{a}=1$. In case (i), $\mathbb{1}\big(b_{nkmp}^{a}=1\big)\pi_{nk}^{mp}=0$, whence there is no impact on the revenues. Moreover, $x_{nka}^{m}=0$. Otherwise, the assignment would lead to a situation with costs but no revenues and therefore suboptimal. Hence, $\mathbb{1}\big(b_{nkmp}^{a}=1\big)\pi_{nk}^{mp}=x_{nka}^{m}y_n^{mp}=0$. In case (ii), whenever $\pi_{nk}^{mp}=0$, we conclude similarly as in case (i). However, if $\pi_{nk}^{mp}=1$, $\mathbb{1}\big(b_{nkmp}^{a}=1\big)\pi_{nk}^{mp}=1$. Also, $x_{nka}^{m}=1$, due to Constraints (3.1g), and $y_n^{mp}=1$, due to Constraints (3.1m). Hence, $\mathbb{1}\big(b_{nkmp}^{a}=1\big)\pi_{nk}^{mp}=x_{nka}^{m}y_n^{mp}$. Unfortunately, a similar reformulation of the total assignment costs (i.e., second term) in Objective (3.1a) is not possible. In this case, we cannot aggregate patterns due to the capacity constraints, as we do with the revenues. This is what distinguishes BA-R from H-A and H-S, as in the latter, they do aggregate scenarios.

Finally, the whole reformulation, called BA-R, can be written as

$$\max \sum_{n\in N}\sum_{k\in K_n}\sum_{m\in M_{nk}}\sum_{p\in P_n^m}\frac{|S_{nk}^{mp}|}{|S|}d_k q_n^{mp}\pi_{nk}^{mp} - \sum_{a\in A}\sum_{m\in M}\sum_{i\in I}\sum_{j\in J}\hat{\mathrm{v}}_a c_{ij}^{m}d_j w_{ija}^{m} - \sum_{i\in I}f_i r_i \tag{3.6}$$

s.t.(3.1b) – (3.1r).

3.3. Computational Results

This section presents a computational study based on Formulation (3.6). We accomplish two main tasks: *First*, we compare the computing times resulting from

the solution of (3.1) using Formulation (3.6) with those obtained under the BA approach of Pinzon Ulloa et al. (2024). *Second*, we supply additional evidence regarding the performance of the solutions produced by H-A and H-S. The section is organized as follows: Section 3.3.1 describes the experimental setup, that is, chiefly, the configurations of the instances involved in each one of the two tasks and the performance metrics. Section 3.3.2 compares the performance of BA-R with that of BA. Section 3.3.3 examines the accuracy of H-A and H-S in comparison with the asymptotically exact BA-R method.

3.3.1. Experimental Setup

This section delineates the main characteristics of our experimental setup. We distinguish two sets of configurations that are aligned with each one of the two tasks at hand. The *first* comprises instances for which an optimal solution can be computed using the BA method (we call them the BA-instances). The *second* set comprises instances that can be solved with the H-A and H-S methods (we call them the H-instances). We define small and large instances. In all experiments, the underlying RUM model is specialized to be a logit. That is, the stochastic elements denoted ε are independently and identically distributed according to the Extreme Value type I distribution with scale parameter β.

Small Instances. The parameter configurations we use for the generation of the small instances are those specified in Section 2.7.2 in Chapter 2, as the instances they yield can be solved with the BA approach. These instances are intentionally small so that they can be solved to optimality even when they are adjoined with large numbers of scenarios. This class of instances features number of facilities $|I| = 2$, number of service levels $|M| = 2$, price level set $P = \{10, 12.5, 15, 17.5, 20\}$, number of shippers $|N| = 1$, number of customers $|J| \in \{15, 30, 45\}$ and number of customer categories $|K| = 3$. Regarding the RUM model, the utility of the opt-out alternative is either one of the values in $U^0 = \{0{,}4.5\}$. We set $\alpha = -0.15$ and consider 11 values for β, thus leading to three classes of instances according to the level of uncertainty.

We use the instance generator from Klose (2018) for u_i, c_{ij}, f_i and d_j, setting the ratio between capacity and demand to $\frac{\sum_{i \in I} u_i}{\sum_{j \in J} d_j} = 1$.

Large Instances. The large instances are those described in Section 2.7.3.3 of Chapter 2. These feature $|N| = 2$, $|K| = 1$, $|M| = 3$ and $|P| \in \{5, 10\}$, whereas each service level is associated with a different set of price levels. We refer to Appendix A.3, Table A.5 for further details. In addition, for the CFLP parameters u_i, c_{ij}, f_i and d_j, we use the ORLib instances presented in Section 7.3.3 of Pinzon Ulloa et al. (2024).

We use the following instances from ORlib: Problems IV, VI, VIII, IX, X. Recall that this set of CFLP instances corresponds to the capacitated warehouse location problem test data sets (used in Beasley, 1988). The RUM parameters are set as follows: $\alpha = -0.1$, $\beta = 1$, $U^0 = 3$ and $L = 4.5$.

Instances and Scenarios. Optimal solutions can be calculated with BA-R for the small instances and $|S| \in \{100, 500, 1000, 5000\}$ and these combinations define the set of B-instances that are examined in Section 3.3.2. The H-instances examined in Section 3.3.3 include the small instances with $|S| = 20{,}000$ as well as the large instances with $|S| \in \{500, 1000, 2000, 4000\}$.

Performance Metrics. Section 3.3.2 reports the ratio between the computing times of BA-R and BA. It also reports the measure of entropy introduced in Legault and Frejinger (2023) so as to evaluate the uncertainty associated with the demand model. It is defined as $\sum_{a \in A} \hat{\mathrm{v}}_a \log(\hat{\mathrm{v}}_a)$. Section 3.3.3 reports the absolute relative gap between the optimal expected profits yielded by methods H-A and BA-R. It is denoted by GAP, and defined as $|EP_{BA-R} - EP_{H-A}|/EP_{BA-R}$, where EP_{H-A} and EP_{BA-R} denote the expected optimal profits respectively obtained with H-A and BA-R.

Computational Setup and Time Budget. We performed the experiments on the Linux version of ILOG CPLEX 12.10 running on an Intel Core i7-7800X CPU at 3.50 GHz. We allocated a 10-hour computing time budget, which is deemed reasonable for the instances examined Section 3.3.3.

3.3.2. Comparison with BA

This section compares the computing times resulting from our proposed BA-R approach with those of the existing BA approach. We only account for the occurrences where the B-instances yield non-zero optimal expected profits in Pinzon Ulloa et al. (2024), since zero expected profits arises from the trivial solution.

Figure 3.1 reports the ratio of the computing time achieved by BA-R in comparison with that of BA when $|S| = 5{,}000$. Figure 3.2 compares ratios for several values of $|S|$. BA-R significantly outperforms BA: on average, the computing time for BA-R is less than 1% of the time required by BA. Whereas the advantage is more noticeable for instances featuring lower entropy values (≤ 2), it is still significant for instances with higher entropy. These results are consistent with the findings of Legault and Frejinger (2023), where instances with low entropy benefit from increased room for aggregation. We also observe that the improvement in computing time with BA-R increases with $|S|$. By using BA-R, we can solve instances with $|S| = 20{,}000$ in less than three minutes whereas an optimal solution could not be calculated with BA in Pinzon Ulloa et al. (2024).

As a further benefit of the speedup afforded by the BA-R reformulation while computing asymptotically exact solutions, it becomes possible to draw additional evidence regarding the accuracies of the heuristic methods H-A and H-S. The next section reports on this.

3.3.3. Comparison with H-A and H-S

This section provides additional information about the accuracies achieved by the H-A and H-S heuristics. It is made possible by the increase in computation speed achieved by the BA-R reformulation in comparison with BA. The use of the considerably slower BA in Pinzon Ulloa et al. (2024) limited the scope of the comparisons with H-A and H-S to simpler instances.

3.3.3.1. Accuracy. Table 3.2 reports the average gaps between BA-R and H-A, and between BA-R and H-S, for the small instances adjoined with the large number of scenarios $|S| = 20{,}000$. It also shows the average gaps between BA and the two

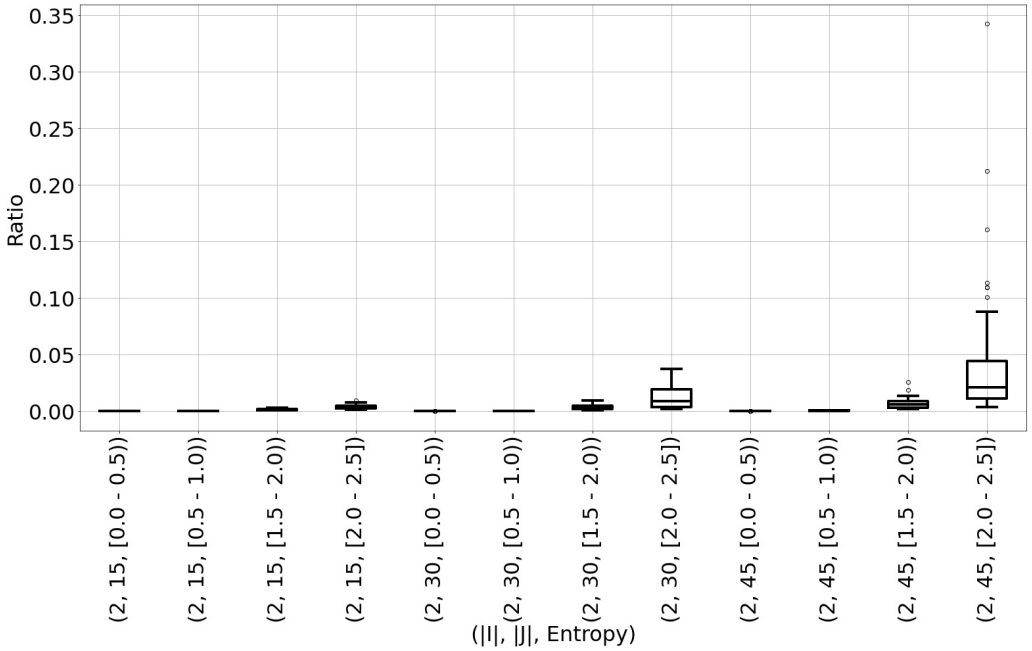


Fig. 3.1. Ratio of computing times with BA-R and BA, given $|I|, |J|$ and entropy in [a,b)

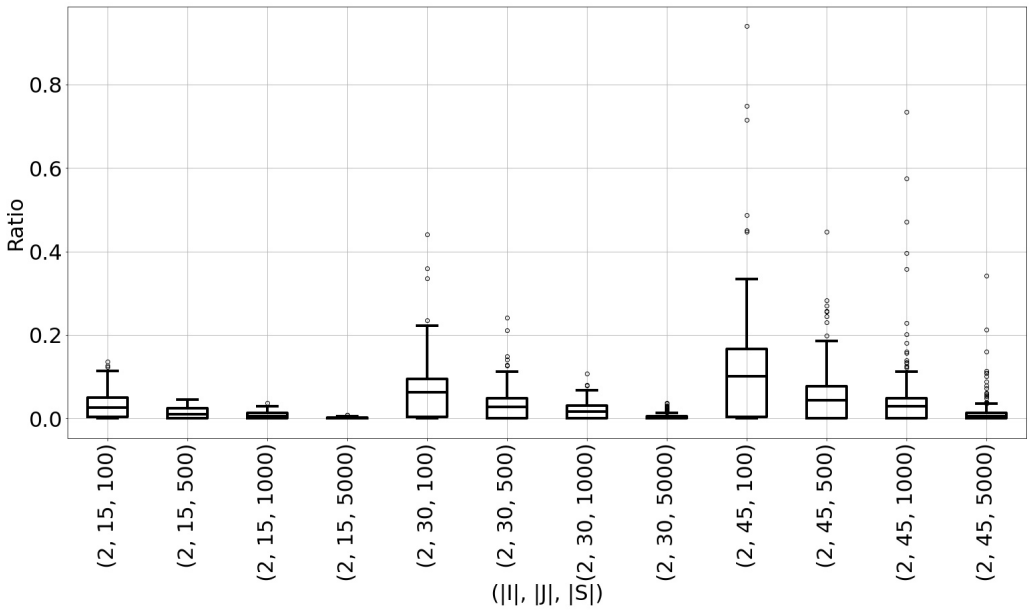


Fig. 3.2. Ratio of computing times with BA-R and BA, given $|I|, |J|$ and $|S|$

heuristics to showcase the gain in accuracy with a larger set of scenarios. Similar results were reported in Pinzon Ulloa et al. (2024), but with $|S| = 5{,}000$. We do not account for occurrences where the average gap was zero for both BA-R and BA. Table 3.3 reports the minimum (min GAP) and maximum (max GAP) gaps between the optimal objective function values obtained by BA-R and H-A for the four sample sizes 500, 1,000, 2,000 and 4,000.

			Average $\lvert GAP \rvert$			
U^0	$\lvert I \rvert$-$\lvert J \rvert$	β	BA-R vs H-S	BA-R vs H-A	BA vs H-S	BA vs H-A
0	2-15	0.5	0.2	0.2	0.1	0.1
0	2-15	1	0.5	0.6	0.6	1.2
0	2-15	2	0.8	1.5	1.5	2.6
0	2-15	4	2.0	3.6	3.0	6.1
0	2-30	0.5	0.1	0.1	0.1	0.1
0	2-30	1	0.4	0.4	0.3	0.6
0	2-30	2	0.6	0.7	0.6	1.2
0	2-30	4	0.8	1.2	0.8	2.1
0	2-30	8	0.8	1.3	1.1	3.3
0	2-30	16	1.0	1.5	1.2	3.9
0	2-30	32	1.0	1.5	1.4	3.9
0	2-45	0.5	0.1	0.1	0.0	0.0
0	2-45	1	0.4	0.4	0.1	0.4
0	2-45	2	0.5	0.5	0.1	0.6
0	2-45	4	0.7	0.7	0.2	1.2
0	2-45	8	0.6	0.6	0.2	1.8
0	2-45	16	0.7	0.7	0.2	2.0
0	2-45	32	0.7	0.7	0.2	2.0
4.5	2-30	16	2.1	2.9	2.5	7.1
4.5	2-30	32	1.5	2.2	1.9	5.2
4.5	2-45	8	1.8	1.8	0.4	4.0
4.5	2-45	16	1.0	1.0	0.3	2.7
4.5	2-45	32	0.8	0.8	0.3	2.3

Table 3.2. Comparison of average GAP obtained with BA-R and BA

Using BA-R, we can solve B-instances with a much larger set of scenarios, in this case we use $|S|$ =20,000. Hence, we can verify that H-A and H-S provide high-quality solutions for these instances, as the average gap with BA-R is significantly reduced in most cases in comparison to BA with $|S| = 5{,}000$. Recall that the BA-R or BA optimal solution asymptotically approximates the optimal solution, whence the average gaps obtained with a larger set of scenarios are more reliable.

Also, using BA-R makes it possible to solve larger problems, such as those of H-instances. As expected, the gap decreases with the larger sets of scenarios due to the closer proximity of the solutions to their asymptotic limits (see Table C.1 in Appendix C.1).

As in in Pinzon Ulloa et al. (2024), the gap between the optimal objective function values obtained by BA-R and H-A is larger when the BA-R optimal solutions are

not stable. We observe that the largest gap obtained for $|S| = 4{,}000$ is lower than 5% for all the cases, and even for instances with $|P| = 10$, where we have not yet attained the desired stability levels with BA-R.

				[min $\|Gap\|$, max $\|Gap\|$]			
Problem set	$\|I\|$-$\|J\|$-$\|P\|$	u_i	f_i	$\|S\| = 500$	$\|S\| = 1{,}000$	$\|S\| = 2{,}000$	$\|S\| = 4{,}000$
IV	16-50-5	5000	7500	[0.9%, 11.1%]	[0.1%, 7.9%]	[0.1%, 4.8%]	[0.6%, 3.2%]
IV	16-50-5	5000	12500	[0.9%, 11.1%]	[0.1%, 7.9%]	[0.1%, 4.8%]	[0.6%, 3.2%]
IV	16-50-5	5000	17500	[0.9%, 11.2%]	[0.1%, 7.9%]	[0.1%, 4.8%]	[0.6%, 3.2%]
IV	16-50-5	5000	25000	[0.9%, 11.2%]	[0.1%, 7.9%]	[0.1%, 4.8%]	[0.6%, 3.2%]
VI	16-50-5	15000	7500	[0.8%, 11.1%]	[0.1%, 7.9%]	[0.1%, 4.8%]	[0.7%, 3.2%]
VI	16-50-5	15000	12500	[0.8%, 11.1%]	[0.1%, 7.9%]	[0.1%, 4.8%]	[0.7%, 3.2%]
VI	16-50-5	15000	17500	[0.8%, 11.2%]	[0.1%, 7.9%]	[0.1%, 4.8%]	[0.7%, 3.2%]
VI	16-50-5	15000	25000	[0.8%, 11.2%]	[0.1%, 7.9%]	[0.1%, 4.8%]	[1.5%, 2.4%]
VII	16-50-5	58268	7500	[0.8%, 11.1%]	[0.1%, 7.9%]	[0.1%, 4.8%]	[0.7%, 3.2%]
VII	16-50-5	58268	12500	[0.8%, 11.1%]	[0.1%, 7.9%]	[0.1%, 4.8%]	[0.7%, 3.2%]
VII	16-50-5	58268	17500	[0.8%, 11.2%]	[0.1%, 7.9%]	[0.1%, 4.8%]	[0.7%, 3.2%]
VII	16-50-5	58268	25000	[0.8%, 11.2%]	[0.1%, 7.9%]	[0.1%, 4.8%]	[0.7%, 3.2%]
IV	16-50-10	5000	7500	[0.6%, 12.3%]	[0.1%, 4.5%]	[1.3%, 5.7%]	[0.3%, 3.9%]
IV	16-50-10	5000	12500	[0.6%, 12.3%]	[0.1%, 4.5%]	[1.3%, 5.7%]	[0.3%, 3.9%]
IV	16-50-10	5000	17500	[0.6%, 12.3%]	[0.1%, 4.5%]	[1.3%, 5.7%]	[0.3%, 3.9%]
IV	16-50-10	5000	25000	[0.6%, 12.3%]	[0.1%, 4.5%]	[1.3%, 5.7%]	[0.3%, 3.9%]
VI	16-50-10	15000	7500	[0.6%, 12.4%]	[0.1%, 4.5%]	[1.3%, 5.7%]	[0.3%, 3.9%]
VI	16-50-10	15000	12500	[0.6%, 12.4%]	[0.1%, 4.5%]	[1.3%, 5.7%]	[0.3%, 3.9%]
VI	16-50-10	15000	17500	[0.6%, 12.4%]	[0.1%, 4.5%]	[1.3%, 5.7%]	[0.3%, 3.9%]
VI	16-50-10	15000	25000	[0.6%, 12.4%]	[0.1%, 4.5%]	[1.3%, 5.7%]	[0.3%, 3.9%]
VII	16-50-10	58268	7500	[0.6%, 12.4%]	[0.1%, 4.5%]	[1.3%, 5.7%]	[0.3%, 3.9%]
VII	16-50-10	58268	12500	[0.6%, 12.4%]	[0.1%, 4.5%]	[1.3%, 5.7%]	[0.3%, 3.9%]
VII	16-50-10	58268	17500	[0.6%, 12.4%]	[0.1%, 4.5%]	[1.3%, 5.7%]	[0.3%, 3.9%]
VII	16-50-10	58268	25000	[0.6%, 12.4%]	[0.1%, 4.5%]	[1.3%, 5.7%]	[0.3%, 3.9%]
VIII	25-50-5	5000	7500	[0.8%, 11.1%]	[0.0%, 7.9%]	[0.2%, 4.8%]	[0.7%, 3.2%]
VIII	25-50-5	5000	12500	[0.8%, 11.1%]	[0.0%, 7.9%]	[0.2%, 4.8%]	[0.7%, 3.2%]
VIII	25-50-5	5000	17500	[0.8%, 11.2%]	[0.0%, 7.9%]	[0.2%, 4.8%]	[0.7%, 3.2%]
VIII	25-50-5	5000	25000	[0.8%, 11.2%]	[0.0%, 7.9%]	[0.2%, 4.9%]	[0.7%, 3.2%]
IX	25-50-5	15000	7500	[0.8%, 11.1%]	[0.0%, 7.9%]	[0.2%, 4.9%]	[0.8%, 3.1%]
IX	25-50-5	15000	12500	[0.8%, 11.1%]	[0.0%, 7.9%]	[0.2%, 4.9%]	[0.8%, 3.1%]
IX	25-50-5	15000	17500	[0.8%, 11.2%]	[0.0%, 7.9%]	[0.2%, 4.9%]	[0.8%, 3.1%]
IX	25-50-5	15000	25000	[0.8%, 11.2%]	[0.0%, 7.9%]	[0.2%, 4.9%]	[0.8%, 3.1%]
X	25-50-5	58268	7500	[0.8%, 11.1%]	[0.0%, 7.9%]	[0.2%, 4.9%]	[0.8%, 3.1%]
X	25-50-5	58268	12500	[0.8%, 11.1%]	[0.0%, 7.9%]	[0.2%, 4.9%]	[0.8%, 3.1%]
X	25-50-5	58268	17500	[0.8%, 11.2%]	[0.0%, 7.9%]	[0.2%, 4.9%]	[0.8%, 3.1%]
X	25-50-5	58268	25000	[0.8%, 11.2%]	[0.0%, 7.9%]	[0.2%, 4.9%]	[0.8%, 3.1%]

Table 3.3. Minimum and Maximum Absolute GAP between BA-R and HA

3.3.3.2. Computing Times. Table 3.4 reports the average computing times when using BA-R for the small instances with $|S| = 20{,}000$. As before, only the instances reporting non-zero profits are accounted for. Table 3.5 reports the average computing times using BA-R for the H-instances. It also shows the computing times when using H-A, as already reported in Pinzon Ulloa et al. (2024), for comparison purposes with

the new results obtained with BA-R. Recall that H-A and H-S solve B-instances in less than 1 second, and that this figure is not affected by the size of S.

Table 3.4 shows that the level of uncertainty impacts the average computing time with BA-R. As expected, with instances presenting a low level of uncertainty ($\beta < 1$), BA-R provides a good performance, comparable to that witnessed with H-A and H-S, due to the higher level of scenario aggregation in such cases. For the large instances, H-A clearly shows better performance than BA-R. The performance of BA-R is affected by the capacity of the facilities: instances with higher capacities are harder to solve, consistent with findings in the literature on capacitated facility location problems (Cornuejols et al., 1991; Klose and Drexl, 2005; Pinzon Ulloa et al., 2024).

Summing up our observations, we are led to the following assessment: By making it possible to compute asymptotically accurate solutions to larger problems than previously possible, BA-R makes it possible to more reliably evaluate the accuracy of the solutions returned by the heuristics H-A and H-S. From the collected evidence that this accuracy is excellent and that H-A and H-S achieve reductions in computation time of several orders of magnitude in comparison with BA and BA-R, we conclude that H-A and H-S are highly attractive candidate methods for solving problems similar to (3.1) in applications.

3.4. Conclusions and Future Directions

This paper has presented a reformulation of a bilevel stochastic facility location and pricing problem with RUM followers, grounded in the scenario aggregation approach proposed in Legault and Frejinger (2023).

The resulting computation method, named BA-R, provides optimal solutions much faster than the original method BA, from which it is derived. We demonstrated that, for instances with low levels of uncertainty, the asymptotically exact BA-R method achieves good computation times, even when the number of scenarios is large, rivaling with those reported for the fast heuristic methods H-A or H-S. By using BA-R instead of BA, we could extend the number of instances (asymptotically) solved

U^0	$\|I\|$-$\|J\|$	β	Avg. computing time (seconds)
0	2-15	0.03125	0.0
0	2-15	0.0625	0.0
0	2-15	0.125	0.0
0	2-15	0.25	0.1
0	2-15	0.5	1.4
0	2-15	1	40.6
0	2-15	2	95.5
0	2-15	4	89.0
0	2-15	8	27.9
0	2-15	16	13.9
0	2-15	32	10.8
0	2-30	0.03125	0.0
0	2-30	0.0625	0.0
0	2-30	0.125	0.0
0	2-30	0.25	0.2
0	2-30	0.5	2.9
0	2-30	1	85.6
0	2-30	2	239.4
0	2-30	4	219.2
0	2-30	8	41.6
0	2-30	16	22.0
0	2-30	32	11.6
0	2-45	0.03125	0.0
0	2-45	0.0625	0.0
0	2-45	0.125	0.0
0	2-45	0.25	0.3
0	2-45	0.5	5.6
0	2-45	1	225.4
0	2-45	2	480.3
0	2-45	4	436.9
0	2-45	8	80.5
0	2-45	16	22.5
0	2-45	32	20.4
0	4-15	0.03125	0.0
0	4-15	0.0625	0.0
0	4-15	0.125	0.0
0	4-15	0.25	0.2
0	4-15	0.5	10.6
4.5	2-30	8	163.3
4.5	2-30	16	40.1
4.5	2-30	32	15.6
4.5	2-45	4	882.1
4.5	2-45	8	127.5
4.5	2-45	16	45.7
4.5	2-45	32	19.1

Table 3.4. Average computing time BA-R, small instances with $|S| = 20{,}000$

to optimality. Hence, we could demonstrate that H-A achieves good accuracies for a wider range of instances than previously possible.

A main outstanding issue is that of solving the embedded CFLP. The CFLP structure impacts the performance of BA-R. Particularly, we witnessed that instances with tight capacities led to a poor performance. Since BA-R embeds $|A|$ CFLP problems (one for each pattern agreement), the impact of the structure of the CFLP problem escalates with the size of A. We believe that an approach that aggregates agreement patterns while leading to a similar CFLP structure can potentially improve

Problem set	$\|I\|$-$\|J\|$-$\|P\|$	u_d	f_d	Avg. computing time BA-R (seconds) ($\|S\| = 4000$)	Computing time H-A (seconds)
IV	16 - 50 -5	5,000	75,00	3,697	6
IV	16 - 50 -5	5,000	12,500	4,136	5
IV	16 - 50 -5	5,000	17,500	3,856	5
IV	16 - 50 -5	5,000	25,000	4,260	5
VI	16 - 50 -5	15,000	7,500	741	4
VI	16 - 50 -5	15,000	12,500	510	4
VI	16 - 50 -5	15,000	17,500	516	5
VI	16 - 50 -5	15,000	25,000	551	6
VII	16 - 50 -5	58,268	7,500	490	6
VII	16 - 50 - 5	58,268	12,500	521	6
VII	16 - 50 - 5	58,268	17,500	485	7
VII	16 - 50 - 5	58,268	25,000	474	6
IV	16 - 50 - 10	5,000	7,500	7,307	12
IV	16 - 50 - 10	5,000	12,500	11,193	12
IV	16 - 50 - 10	5,000	17,500	13,288	13
IV	16 - 50 - 10	5,000	25,000	7926	13
VI	16 - 50 - 10	15,000	7,500	676	10
VI	16 - 50 - 10	15,000	12,500	885	10
VI	16 - 50 - 10	15,000	17,500	594	10
VI	16 - 50 - 10	15,000	25,000	1,006	11
VII	16 - 50 - 10	58,268	7,500	1,311	17
VII	16 - 50 - 10	58,268	12,500	566	15
VII	16 - 50 - 10	58,268	17,500	1,331	18
VII	16 - 50 - 10	58,268	25,000	2,048	13
VIII	25 - 50 - 5	5,000	7,500	8,980	13
VIII	25 - 50 - 5	5,000	12,500	10,308	14
VIII	25 - 50 - 5	5,000	17,500	14,958	13
VIII	25 - 50 - 5	5,000	25,000	12,045	14
IX	25 - 50 - 5	15,000	7,500	679	13
IX	25 - 50 - 5	15,000	12,500	677	11
IX	25 - 50 - 5	15,000	17,500	677	19
IX	25 - 50 - 5	15,000	25,000	702	21
X	25 - 50 - 5	58,268	7,500	646	13
X	25 - 50 - 5	58,268	12,500	670	12
X	25 - 50 - 5	58,268	17,500	674	22
X	25 - 50 - 5	58,268	25,000	718	15

Table 3.5. Comparison of computing times BA-R vs H-A, H-instances

computing times, as this will avoid evaluating feasible solutions repetitively. This issue will be addressed in future work.

Chapter 4 Third article

A Capacitated Collection-and-Delivery-Point Location Problem with Random Utility Maximizing Customers

by

David Pinzon Ulloa[1], Ammar Metnani[1], and Emma Frejinger[1]

([1]) Département d'informatique et de recherche opérationnelle, Université de Montréal, Montréal, Canada.
CIRRELT, Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation

This article has been submitted to Computers and Operations Research.

The main contributions of the authors for this article are as follows.

- David Pinzon proposed the model formulations, designed, implemented, ran the experiments, and wrote the article.
- Ammar Metnani contributed to the mathematical formulations and writing of the article.
- Emma Frejinger provided advice on the methodology and contributed to the mathematical formulations and writing of the article.

ABSTRACT. We consider a strategic decision-making problem where a logistics provider (LP) seeks to locate collection and delivery points (CDPs) with the objective to reduce total logistics costs. The customers maximize utility that depends on their perception of home delivery service as well as the characteristics of the CDPs, including their location. At the strategic planning level, the LP does not have complete information about customers' preferences and their exact location. We introduce a mixed integer non-linear formulation of the problem and propose two linear reformulations. The latter involve sample average approximations and closest assignment constraints, and in one of the formulations we use scenario aggregation to reduce its size. We solve the formulations with a general-purpose solver using a standard Benders decomposition method. Based on extensive computational results and a realistic case study, we find that the problem can be solved efficiently. However, the level of uncertainty in the instances determines which approach is the most efficient. We use an entropy measure to capture the level of uncertainty that can be computed prior to solving. Furthermore, the results highlight the value of accurate demand modeling, as customer preferences have an important impact on the solutions and associated costs.
Keywords: Capacitated facility location, parcel lockers, closest assignment constraints, random utility maximization models

4.1. Introduction

The sustained increase in e-commerce transactions and online sales has promoted a rapid and sharp increase in demand for last-mile deliveries (Rohmer and Gendron, 2020). Moreover, customers are demanding fast deliveries with high service levels and expect flexible shipping options (Amorim et al., 2024). This imposes increasing

pressure on the distribution network of logistics providers (LP). Moreover, it may negatively impact the environment by increasing emissions due to failed and repeated deliveries, as well as contributing to increased congestion (Molin et al., 2022). Alternative delivery options, in particular, collection and delivery points (CDP), arise as means of reducing costs and emissions (Yuen et al., 2018).

Using CDPs as an alternative to home delivery is not uncommon and its adoption varies around the world. For example, in Singapore, 5.5% of online shoppers have adopted parcel lockers as a delivery option (Choo, 2016), while this figure is 20% in France and 10% in Germany (Morganti et al., 2014). The main impacts of CDPs are (i) reducing delivery costs thanks to improved consolidation and (ii) reduced return levels due to better first-delivery rates. Some customers may also perceive it as an improved service level.

In this work, we address a strategic decision-making problem faced by a LP seeking to reduce costs by locating different types of CDPs to divert demand from its home delivery service. The problem crucially depends on the demand for services via CDPs, that is, it depends on customers' preferences. In this context, two sources of uncertainty arise. *First*, at a strategic planning level, the LP does not have perfect knowledge about its future customers and their precise locations. *Second*, the LP has imperfect information about customers' preferences. We assume that the LP has access to historical data capturing past customers' characteristics and preferences, and that such data can be used to estimate a random utility maximization (RUM) model to predict demand (McFadden, 1981). We propose to model our problem with a mixed integer non-linear formulation that integrates a RUM model of customer behavior.

Our work is related to the growing body of literature on facility location problems where user's behavior is described with RUM models, also known as choice-based facility location problems. Similar to several studies (e.g., Pacheco Paneque et al., 2021; Pinzon Ulloa et al., 2024; Haase and Müller, 2014; Legault and Frejinger, 2023), we use a sample average approximation (SAA) to derive a linear formulation for any additive RUM model. Closest to our work are the studies addressing facility location with fixed capacities (Haase and Müller, 2013; Pinzon Ulloa et al., 2024). Also

relevant to our work is the literature on location routing problems (LRPs), and, in particular, studies where not all customers are assigned to a delivery route and where they can get serviced by visiting a facility (Arslan, 2021; Stenger et al., 2012). In this context, only a few studies consider customers' preferences (Janjevic et al., 2019; Guerrero-Lorente et al., 2020) and, unlike our model, they consider the preferences to be perfectly known (i.e., described by a deterministic demand model). Similar to some works on LRPs (Janjevic et al., 2019; Guerrero-Lorente et al., 2020), we do not solve a routing problem due to its computational difficulties. Instead, we use routing cost estimations, in our case based on average routing costs from historical data.

We contribute to the literature by addressing a capacitated CDP location problem where there is uncertainty about customers' locations and preferences. In contrast to the closely related work of Haase and Müller (2013) and Pinzon Ulloa et al. (2024), we tackle two challenging aspects: (i) potentially conflicting objectives between the LP and its customers, and (ii) strict assignments of customers to facilities. The main methodological contribution of our work is the introduction of formulations that can be effectively solved using a standard Benders decomposition method implemented in general-purpose solvers. This is achieved through a specific formulation of the follower (customer) problem in a bilevel formulation based on SAA. We demonstrate how this follower problem leads to the well-known *closest assignment constraints* in our single-level linear reformulation. Furthermore, we present an extensive computational study analyzing the computing times and the effects of uncertainty on the solutions. Similar to Legault and Frejinger (2023), we introduce an entropy measure and show how computational performance varies with the level of uncertainty in the instances. Finally, we showcase the practical applicability of our work by detailing a realistic case study of our industrial partner.

The remainder of this paper is structured as follows: Section 4.2 reviews related work and Section 4.3 describes the problem we address in this work. Section 4.4 introduces the non-linear mathematical model that we propose, and Section 4.5 details an illustrative example. In Section 4.6, we derive mixed integer linear programming (MILP) reformulations and discuss the closest assignment constraints. Section 4.7

reports the computational study and Section 4.8 presents the case study. Section 4.9 concludes.

4.2. Literature Review

This section reviews the literature related to our work. It mainly concerns facility location problems integrating choice models. However, we briefly comment on the literature related to LRPs as our location problem is impacted by routing costs, although we do not model that problem in detail. Finally, we overview the main contributions of our work.

There is a growing body of literature on facility location problems including RUM models, also known as choice-based facility location problems (Hui Lin et al., 2024). We distinguish the works that address the problem using closed-form probabilities (e.g., Ljubić and Moreno, 2018; Mai and Lodi, 2020; Krohn et al., 2021; Duong et al., 2023) from those that address them by simulating utilities (e.g., Haase and Müller, 2013; Pacheco Paneque et al., 2021; Legault and Frejinger, 2023; Lamontagne et al., 2023; Pinzon Ulloa et al., 2024). Our work resides within the latter group but, unlike most studies, we consider capacitated facilities. We are only aware of two works (Haase and Müller, 2013; Pinzon Ulloa et al., 2024) that use simulated utilities and treat capacitated facilities (note that, e.g., Pacheco Paneque et al., 2021, also consider capacities but for a different problem setting). Our work differs from Haase and Müller (2013) in three main aspects. *First*, in their case, the open facilities must satisfy all of the demand, while we consider an outside option. *Second*, we consider a bilevel formulation because the objective of the LP and the customers are not necessarily aligned. In their case, the facility location objective function is aligned with the objective of the users as both maximize utility. *Third*, they do not make strict customer assignments to facilities, whereas we do. Pinzon Ulloa et al. (2024) consider a pricing and location problem where the random utilities are associated with price and service levels. However, these utilities do not explicitly depend on facility location decisions. Moreover, similar to Haase and Müller (2013), Pinzon Ulloa et al. (2024) do not make strict customer assignments to facilities.

Other works have addressed competitive facility location problems (see, e.g., Beresnev and Melnikov, 2016, 2018; Casas-Ramirez et al., 2018) and capture problems (see e.g., Berman et al., 2007) with capacities. These works use deterministic patronizing rules instead of RUM models.

In regard to applications, the problem of locating parcel lockers is receiving increasing attention (see, Rohmer and Gendron, 2020, for a review). We identify two classes of related work. The *first* class assumes perfect knowledge of customer preferences (Deutsch and Golany, 2018; Luo et al., 2022). That is, the demand models are deterministic. On the contrary, the *second* class involves models that do not assume perfect knowledge about demand and does not use closed-form stochastic demand models (e.g., Lin et al., 2020, 2022). Our work defines a new class where the simulation approach opens the formulation to any type of RUM model.

Our work seeks to locate facilities so as to reduce estimated routing costs. However, we do not explicitly model the routing problem as is done in the LRP literature. This is a rich area of research (see e.g., Mara et al., 2021; Lopes et al., 2013; Prodhon and Prins, 2014) covering several problem variants (e.g. Drexl and Schneider, 2015). Thus, several studies on LRP consider the location of depots that constitute supply points with the objective of serving every customer with delivery tours starting at one of the depots (see e.g., Nagy and Salhi, 2007; Schiffer and Walther, 2018; Arslan, 2021). There are also problem variants closer to our setting, such as location or routing problems (see Arslan, 2021) and LRP with subcontracting options (see Stenger et al., 2012). However, there is limited work on problem settings with parcel lockers (see e.g., Zhou et al., 2016; Janjevic et al., 2019; Guerrero-Lorente et al., 2020). Only Janjevic et al. (2019) and Guerrero-Lorente et al. (2020) consider customer's preferences for the pick-up locations and both studies assume perfect knowledge of customers' preferences. Moreover, they focus on challenges associated with approximating routing costs. Instead, we focus on modeling demand uncertainty, and we estimate routing cost using historical data.

Table 4.1 summarizes the key aspects of our work in comparison to the literature. We consider a CDP location problem with capacitated facilities and a stochastic demand model where the customers' utility-maximizing objective may be conflicting

with the facility location objective. Uncertainty arises because, at the strategic planning horizon, customer locations are not perfectly known. Moreover, we assume imperfect knowledge of customers' preferences which we capture with an additive RUM model without making any specific assumptions about its type. The next section introduces the problem in greater detail.

Reference	Parcel Lockers	Choice Model	Capacity	Strict Assignments	Bilevel program	Conflicting objectives
Berman et al. (2007)		D	✓	✓		
Beresnev and Melnikov (2016)		D	✓	✓	✓	✓
Beresnev and Melnikov (2018)		D	✓	✓	✓	✓
Casas-Ramirez et al. (2018)		D			✓	✓
Deutsch and Golany (2018)	✓	D				
Duong et al. (2023)		MNL				
Guerrero-Lorente et al. (2020)	✓	D				
Haase and Müller (2013)		Any	✓			
Janjevic et al. (2019)	✓	D				
Krohn et al. (2021)		MNL				
Lamontagne et al. (2023)		Any			✓	
Lin et al. (2020)	✓	MNL				
Lin et al. (2022)	✓	TLM				
Ljubić and Moreno (2018)		MNL				
Luo et al. (2022)	✓	D				
Mai and Lodi (2020)		MNL				
Méndez-Vogel et al. (2023)		PBL			✓	✓
Legault and Frejinger (2023)		Any			✓	
Pinzon Ulloa et al. (2024)		Any	✓		✓	✓
Zhou et al. (2016)	✓					
Our work	✓	Any	✓	✓	✓	✓

D: Deterministic
MNL: Multinomial Logit Model
TLM: Threshold Luce Model
PBL: Partially Binary Logit Model

Table 4.1. Summary of related work

4.3. Problem Description

This section details our strategic decision-making problem that consists of determining the locations for CDPs – drop-off and pick-up facilities – to minimize routing, operating, and implementing costs resulting from such a decision over a planning period. We begin by introducing the types of facilities involved, then discuss the specific characteristics of the demand, and conclude with the impact of routing costs.

The LP provides a parcel delivery service directly to its customers' home locations (standard home delivery). As a complement, it seeks to implement additional delivery services using CDPs. These CDPs are alternative locations where customers can pick up or drop off their parcels (see, e.g. Janjevic et al., 2019; Rohmer and Gendron, 2020). We distinguish two main types of CDPs: pick-up (P) points, often integrated into stores or local shops, and automated parcel lockers, which do not require staff assistance during pick-up or drop-off. The LP considers two types of parcel lockers: Regular (A) and modular (M). Regular parcel lockers have fixed capacity, whereas in modular parcel lockers, the capacity can increase by multiples of a base capacity. Let D denote the set of candidate locations for CDPs. Also, let D^{P}, D^{A} and D^{M} denote the corresponding sets for pick-up points P, and parcel lockers A and M, respectively. Each $d \in D$ has coordinates $\theta_d = (\theta_d^{\mathrm{x}}, \theta_d^{\mathrm{y}}) \in \mathbb{R}^2$. Moreover, $d \in D^{\mathrm{P}} \cup D^{\mathrm{A}}$ has a fixed capacity u_d, whereas deciding the capacity of $d \in D^{\mathrm{M}}$ is part of our problem.

The LP does not have precise knowledge about future customers and deliveries. Customers are, therefore, aggregated based on their geographical location and preferences. More precisely, customers are classified into categories $k \in K$ according to delivery profiles. These delivery profiles are based on customers' characteristics and preferences captured in historical data. Furthermore, we consider two levels of geographical aggregation. CDPs attract demand from customers in a zone z within a set of zones Z. To model demand, each zone is further partitioned into subzones $a \in A_z$. Customers are distributed across subzones, each having an expected yearly number of parcels b_{zak}, $z \in Z, a \in A_z, k \in K$ with a total number of parcels within a zone $b_z = \sum_{a \in A_z} \sum_{k \in K} b_{zak}$. Note that prediction of these parameters is harder when more detailed data is necessary.

The demand that each CDP attracts is influenced by the customers' preferences for this option relative to the standard home delivery service. The distance from the customer's location to a CDP is an important explanatory factor. Given the geographical aggregation of customers, we have some distance information (in the experimental section, we consider the distance from the centroid of the corresponding subzone a, denoted by $\theta_a \in \mathbb{R}^2$). Other factors can impact the customers' preferences and these are not all known to the LP. This motivates the use of a stochastic demand

model in our case. Note that, as the demand captured by a CDP cannot exceed its capacity, excess demand is reallocated to the home delivery service.

Implementing CDPs implies variable (i.e., volume-dependent) and fixed costs $f_d > 0$, $d \in D^{\mathrm{A}} \cup D^{\mathrm{M}}$. Note that fixed costs increase with capacity for $d \in D^{\mathrm{M}}$. Moreover, typically $f_d = 0$, $d \in D^{\mathrm{P}}$, but these pick-up points require a minimum demand level $\underline{b}_d$ from the LP to respect business agreements with the stores.

Variable costs are mainly routing costs to serve the customers. The LP estimates the current unit routing costs per parcel based on historical delivery and routing operations for each zone, c_z^{R}, $z \in Z$. They encompass both transit costs, incurred when the vehicle moves between a terminal to the corresponding zones and within the zone, as well as delivering-operating costs related to loading, sorting, and moving parcels from the vehicle to the corresponding delivery addresses.

The use of CDPs impacts the estimated routing costs as fewer customers require home delivery services, and the consolidated demand from captured customers requires less transit and delivering-operating costs. In this problem, explicit routing decisions cannot be made as the LP uses customer data aggregated over subzones. However, the LP considers that there is an impact on the delivering-operating costs due to parcel consolidation when delivering to a CDP instead of to different home addresses. Hence, the LP quantifies the affected routing (AR) costs of customers in a zone z served by a CDP d by assuming a reduction in the delivering-operating costs, regardless of the transit costs. We denote these costs by c_{dz}^{AR}. Note that in this case $c_{dz}^{\mathrm{AR}} \leq c_z^{\mathrm{R}}$.

In summary, the LP seeks to optimally locate CDPs to minimize total expected costs encompassing estimated routing costs and CDP's fixed and operational costs. Since these costs depend on the demand for CDPs, the LP must consider customers' preferences for CDPs, their capacity as well as the preferences for home delivery services.

4.4. A Mixed Integer Non-Linear Formulation

This section introduces a non-linear formulation for our problem. It is the foundation of the two linear reformulations that are derived and discussed in Section 4.6. We begin by presenting the main modeling concepts and then we focus on the mathematical formulation itself.

We denote by x_d the binary decision of implementing or not a CDP at $d \in D$ (in vector notation $\boldsymbol{x} \in \{0,1\}^{|D|}$). We introduce demand models at the two levels of geographical aggregation (zone and subzone) that depend on $\boldsymbol{x}$.

The share of customers in z choosing delivery service d is denoted $\psi_{dz}(\boldsymbol{x})$. Since the captured demand is constrained by the CDP's capacity, we define the effective share of demand captured by a CDP d in zone z and denoted it with the auxiliary variables p_{dz}. Note that the latter implicitly depends on $\boldsymbol{x}$ since $p_{dz} \leq \psi_{dz}(\boldsymbol{x})$.

Customer information is available at the subzone level. We model demand for CDPs and home delivery services with a RUM model and we denote the probability of customers in category k and subzone a choosing d by $\psi_{dak}(\boldsymbol{x})$. Accordingly, the demand shares, at zone level, are defined by

$$\psi_{dz}(\boldsymbol{x}) = \sum_{a \in A_z} \sum_{k \in K} \frac{b_{zak}}{b_z} \psi_{dak}(\boldsymbol{x}), \tag{4.1}$$

with

$$\psi_{dak}(\boldsymbol{x}) = \text{Prob}\left(d \in \underset{d' \in D_x \cup \{0\}}{\arg\max} \{U_{d'ak}(\boldsymbol{\varepsilon_{d'}})\} \right). \tag{4.2}$$

In (4.2), $U_{d'ak}(\boldsymbol{\varepsilon_{d'}})$ is an additive random utility function capturing the attractiveness of a CDP d' for customers of category k in subzone a. It is defined over the set of open CDPs $D_{\boldsymbol{x}}$ and home delivery service (indexed with 0). For simplicity, we henceforth write U_{dak} instead of $U_{d'ak}$ and define it as

$$U_{dak} = \alpha_k^1 f^1(\theta_d, \theta_a) + \alpha_k^2 f^2(E_d) + \varepsilon_{dak}, \tag{4.3}$$

where $f^1(\theta_d, \theta_a)$ denotes a function of the distance from the customers located in subzone $a \in A_z$ to the location of the CDP d, endogenous to our model as it depends on decision variables x_d, and $f^2(E_d)$ is a function of other features, exogenous to

our model, that may explain customer preferences. Parameters α_k^1 attached to CDP alternatives are strictly negative and reflect that the probability of choosing a CDP decreases as distance increases. For home delivery service, $\alpha_k^1 = 0$. Parameters $\alpha_k^2 \geq 0$ reflect the relative importance of other attributes for customer category k. Terms ε_{dak} are continuous random variables with support Ξ. Specific distributional assumptions lead to different types of RUM models. For example, assuming that ε_{dak} are independently and identically distributed extreme value type I leads to the well-known logit model.

Modular parcel lockers $d \in D^{\mathrm{M}}$ have capacities that increase by multiples of a base capacity. Accordingly, we define a set of capacity levels $L_d = \{u_d^1, u_d^2, \ldots, u_d^{\mathrm{L}}\}$, each with a corresponding fixed cost denoted by f_d^l. Moreover, we introduce binary decision variables r_d^l that equal one if level l is selected for CDP d, and zero otherwise.

The mixed integer non-linear formulation of our problem is:

$$\min \sum_{z \in Z} c_z^{\mathrm{R}} b_z p_{0z} + \sum_{z \in Z} \sum_{d \in D} c_{dz}^{\mathrm{AR}} b_z p_{dz} + \sum_{d \in D^{\mathrm{M}}} \sum_{l \in L_d} f_d^l r_d^l + \sum_{d \in D^{\mathrm{A}}} f_d x_d \tag{4.4a}$$

$$\text{s.t.} \sum_{l \in L_d} r_d^l = x_d, \qquad d \in D^{\mathrm{M}}, \tag{4.4b}$$

$$\sum_{z \in Z} b_z p_{dz} \leq \sum_{l \in L_d} u_d^l r_d^l, \qquad d \in D^{\mathrm{M}}, \tag{4.4c}$$

$$\sum_{z \in Z} b_z p_{dz} \leq u_d x_d, \qquad d \in D^{\mathrm{P}} \cup D^{\mathrm{A}}, \tag{4.4d}$$

$$\sum_{z \in Z} b_z p_{dz} \geq \underline{b}_d x_d, \qquad d \in D^P, \tag{4.4e}$$

$$p_{dz} \leq \psi_{dz}(\boldsymbol{x}), \qquad z \in Z, d \in D, \tag{4.4f}$$

$$p_{0z} \geq \psi_{0z}(\boldsymbol{x}), \qquad z \in Z, \tag{4.4g}$$

$$p_{0z} + \sum_{d \in D} p_{dz} = 1, \qquad z \in Z, \tag{4.4h}$$

$$x_d \in \{0,1\}, \qquad d \in D, \tag{4.4i}$$

$$p_{dz} \geq 0, \qquad z \in Z, d \in D \cup 0, \tag{4.4j}$$

$$r_d^l \in \{0,1\}, \qquad d \in D^{\mathrm{M}}, l \in L_d. \tag{4.4k}$$

Objective function (4.4a) minimizes the total estimated routing costs, including the costs of implementing and operating CDPs. It has four components capturing the (i) total estimated routing costs for home-delivery service, (ii) total estimated routing and operating costs related to demand allocated to open CDPs, as well as total fixed costs of implementing (iii) modular and (iv) regular parcel lockers, respectively. Constraints (4.4b) impose that only one of the available capacity levels can be chosen for open modular parcel lockers. Constraints (4.4c) and (4.4d) ensure that the expected demand captured by a CDP does not exceed its capacity, whereas Constraints (4.4e) impose a minimum demand level for pickup points $d \in D^{\mathrm{P}}$. Constraints (4.4f) impose that the effective fraction of demand captured by a CDP cannot exceed the demand share. In contrast, Constraints (4.4g) state that the effective fraction of demand captured by the home-delivery service must be at least equal to the demand share. We include these constraints for the sake of clarity but note that they are, in fact, redundant.

Constraints (4.4h) ensure that all the demand is captured by either a CDPs or the home delivery service.

We note that model (4.4) is non-linear due to Constraints (4.4f), in view of the definitions of $\psi_{dz}(\boldsymbol{x})$ in (4.1) and of the individual probabilities in (4.2). Indeed, even the simplest RUM model – multinomial logit – is non-linear. For more advanced RUM models, such as the logit mixtures (Train, 2009), (4.2) does not even have a closed form whence simulation is required to evaluate probabilities. In this situation, one way to solve model (4.4), is to introduce simulated utilities that are maximized as part of a follower problem in a bilevel formulation (e.g. Pacheco Paneque et al., 2021; Lamontagne et al., 2023; Pinzon Ulloa et al., 2024). In our case, the bilevel structure is important, as the RUM objective of the user may not be aligned with the cost minimization objective of the LP. We describe such reformulations in Section 4.6 after the following example designed to illustrate the impact of capacities and costs

on the location decisions and captured demand. Finally, we refer to Tables B.1 and B.2 in Appendix B.1 for a comprehensive listing of the notation.

4.5. Illustrative Example

This section illustrates the impact of locating CDPs in a region divided into two zones, where customers are served by the LP's standard home delivery service. We discuss how capacity constraints affect the demand captured by these facilities and highlight the balance the LP must achieve between capturing demand and minimizing costs.

We consider two zones z_1 and z_2 and two candidate locations for CDPs $\{A,B\}$, as depicted in Figure 4.1. In this example, the demand captured by the location A is expected to be greater than in location B. For instance, consider A to be a popular convenience store, which makes it a more attractive option for customers to pick up and deliver orders compared to B, which has a less attractive location.

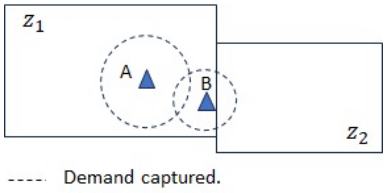


Fig. 4.1. Zones and CDPs

$\psi(\boldsymbol{x})$	$x_A = 1, x_B = 0$	$x_A = 0, x_B = 1$	$x_A = x_B = 1$
ψ_{Az_1}	0.35	0	0.30
ψ_{Bz_1}	0	0.20	0.15
ψ_{0z_1}	0.65	0.80	0.55
ψ_{Az_2}	0	0	0
ψ_{Bz_2}	0	0.05	0.05
ψ_{0z_2}	1	0.95	0.95

Table 4.2. Choice probabilities for each CDP location decision

The demand in each zone is given $b_{z_1} = b_{z_2} = 500$, respectively. The unit costs of home delivery service in each zone are $c^{\mathrm{R}}_{z_1} = 5$ and $c^{\mathrm{R}}_{z_2} = 6$. For the sake of simplicity, we assume a single customer category and no partition into subzones. The capacity

for the CDP in each location is given by $u_A = 90, u_B = 130$, and the fixed costs are $f_A = 170, f_B = 200$.

Figure 4.1 shows the relative location of CDPs and highlights that, from a distance perspective, they are more attractive for customers in zone z_1. We fix the choice probabilities as reported in Table 4.2 for different location decisions x_d. Note that these probabilities are fixed following our assumption that a CDP located in a store (CDP A) attracts more demand than that located in (CDP B). Recall that the effective fraction captured by a CDP can be lower than the corresponding probabilities ψ, due to the capacity constraints.

The estimated costs of serving customers captured by the CDPs are $c_{Az_1} = c_{Az_2} = 3.25$ and $c_{Bz_1} = c_{Bz_2} = 3.5$. If none of the CDPs are open, the total estimated routing cost is \$5,500 $(5 \cdot 500 + 6 \cdot 500)$. This corresponds to serving all the customers with home-delivery services, thus $p_{0z_1} = p_{0z_2} = 1$. However, the optimal solution is $x_A = 0, x_B = 1$, with a total cost of \$5,488. In this solution, CDP B captures 20% of the demand from zone A and 5% from zone B. Even though CDP A is more attractive and less expensive than CDP B, its capacity limits the amount of demand captured, impacting the potential cost reduction. We note that the optimal solution is sensitive to the capacities of CDPs. For example, if the capacity of CDP A increases by at least three units, then the optimal solution changes to $x_A = 1, x_B = 0$. Table 4.3 summarizes the results.

	$x_A = 1$ $x_B = 0$	$x_A = 0$ $x_B = 1$	$x_A = x_B = 1$	$x_A = x_B = 0$
Total cost	5,513	5,488	5,538	5,500
Captured demand zone z_1	90	100	165	0
Effective fraction zone z_1	18%	20%	33%	0
Captured demand zone z_2	0	25	25	0
Effective fraction zone z_2	0	5%	5%	0

Table 4.3. Results for location decisions x_d

The results in Table 4.3 also highlight the impact of the conflicting objectives between customers and the LP. We observe that a solution capturing more demand (more customers allocated to their preferred choice) is not necessarily the best for the LP. Indeed, when both CDPs are open, they capture more demand. However, this solution is more costly than using only home delivery services for all the customers.

Recall the rule that demand is allocated to home delivery service in case of excess demand. When both CDPs are open, the effective fraction of demand captured follows the LP's allocation rule. This fraction results in a lower bound on the effective fraction that both CDPs can capture when customers can be reallocated to the CDP next in preference. To show this, let us consider that the excess of demand captured by CDP A is reallocated to CDP B until its maximum capacity is reached. Then, the total captured demand from zone z_1 would be 195. That corresponds to 39% ($> 33\%$) of the demand in zone z_1 and represents a lower cost (\$5,493) than the one displayed in Table 4.3 for the same solution ($x_A = x_B = 1$). However, in this case, the corresponding total cost with this allocation rule is still higher than the cost obtained by the optimal solution under the allocation assumption in this work.

4.6. Simulation-based Formulations

This section presents our approaches to deal with the non-linear program (4.4) introduced in Section 4.4. They are bilevel MILP formulations based on SAA of customer choice probabilities. We begin this section by discussing the foundation of the SAA formulation and two modeling alternatives for the follower's problem. Then, we focus on the modeling alternative leading to the single-level formulation with the smallest number of variables and constraints. We introduce in Section 4.6.1 the single-level reformulation, and then, in Section 4.6.2, an equivalent reformulation based on scenario aggregation.

We adopt an existing approach that consists in simulating realizations of random utilities that are linear in decision variables (e.g. Haase and Müller, 2013; Pacheco Paneque et al., 2021). We draw a sample set S of scenarios $\{\boldsymbol{\varepsilon_s}\}_{s \in S}$ from the distribution of the random terms ε_{dak} in (4.3). We estimate $\psi_{dak}(\boldsymbol{x})$ with an SAA through

$$\hat{\psi}_{dak}(\boldsymbol{x}) = \frac{1}{|S|} \sum_{s \in S} w_{daks}, \qquad d \in D, a \in A, k \in K \tag{4.5a}$$

$$w_{daks} = \begin{cases} 1, & d \in \underset{d' \in D_x \cup \{0\}}{\arg\max} \{U_{d'aks}\} \\ 0, & \text{otherwise.} \end{cases} \qquad d \in D, a \in A, k \in K, s \in S \tag{4.5b}$$

In (4.5b), w_{daks} equals one if delivery service d (a CDP or home-delivery service) has the maximum utility in scenario s for customers in category k and subzone a. These are decision variables in the follower problem of our bilevel formulation. We consider two formulations: For given a, k, and s, and (leader) decision variables $\boldsymbol{x}$, a first, straightforward formulation of (4.5b) is

$$\arg\max \sum_{d \in D} U_{daks} x_d w_{daks} + U_{0aks} w_{0aks} \tag{4.6a}$$

$$\text{s.t.} \sum_{d \in D \cup \{0\}} w_{daks} = 1, \tag{4.6b}$$

$$w_{daks} \in \{0,1\}, \qquad d \in D \cup \{0\}. \tag{4.6c}$$

and a second, alternative formulation is

$$\arg\max \sum_{d \in D \cup \{0\}} U_{daks} w_{daks} \tag{4.7a}$$

$$\text{s.t.} \sum_{d \in D \cup \{0\}} w_{daks} = 1, \tag{4.7b}$$

$$w_{daks} \leq x_d, \qquad d \in D, \tag{4.7c}$$

$$w_{daks} \in \{0,1\}, \qquad d \in D \cup \{0\}. \tag{4.7d}$$

We note that the coefficient matrices in these two formulations are totally unimodular. Therefore, we can relax the integrality constraints on variables w_{daks} and still obtain binary solutions. We also note that without loss of generality, we assume that $U_{daks} > 0$. Indeed, in the case of negative utilities, we can add a constant (e.g., $\hat{U}_{daks} = U_{daks} - \delta$, where $\delta < \min_{d' \in D \cup \{0\}} U_{d'aks}$) without affecting the solutions. Based on this assumption, and the fact that U_{daks} are sampled from continuous distributions (the probability that any two values are equal is zero), the optimal solutions to the formulations are unique. In addition, in the following proposition, we show that the optimal solutions of the two formulations are equal. euiue

Proposition 4. *For given a, k, s, let $\hat{\boldsymbol{w}}_{aks}$ and $\tilde{\boldsymbol{w}}_{aks}$ be the vectors representing the corresponding optimal solutions for* (4.6) *and* (4.7) *given a set of open facilities $D^* \subset D$ defined by $\boldsymbol{x}$. Then $\hat{\boldsymbol{w}}_{aks} = \tilde{\boldsymbol{w}}_{aks}$.*

Proof. By inspection of (4.6) and (4.7), we see that any level of utility reachable in the objective of (4.6) can also be reached in (4.7) by adjusting $\tilde{\boldsymbol{w}}_{aks}$ and conversely that any level of utility reachable in (4.7) can also be reached in (4.6) by adjusting $\hat{\boldsymbol{w}}_{aks}$. Hence the sets of utility levels reachable in (4.6) and (4.7) are identical. For simplicity, we henceforth omit indices a,k, and s. We have $x_d = 1, d \in D^*$. Let us assume an optimal solution to (4.6): $\hat{w}_{d'} = 1, d' \in D^* \cup \{0\}$, and $\hat{w}_d = 0, d \neq d' \in D^* \cup \{0\}$. This implies that $U_{d'} > U_d, d \neq d' \in D^* \cup \{0\}$. Let us assume a different optimal solution to (4.7): $\tilde{w}_{d''} = 1, d'' \neq d' \in D^* \cup \{0\}$, and $\tilde{w}_d = 0, d \neq d'' \in D^* \cup \{0\}$. This implies that $U_{d''} > U_d, d \neq d'' \in D^* \cup \{0\}$. However, this contradicts the assumption that $\hat{\boldsymbol{w}}$ is optimal for (4.6). Hence, both optimal solutions must be equal. □

In the following section, we introduce a single-level formulation using (4.7) where a subset of the constraints are closest assignment constraints (CAC). In Section 4.6.2, we introduce another single-level formulation that is also based on (4.7), but that uses an aggregation of scenarios inspired by Legault and Frejinger (2023).

4.6.1. CAC Formulation

Next, we introduce a linear single-level formulation of (4.4). It is derived by replacing $\psi_{dz}(\boldsymbol{x})$ in Constraints (4.4f) by their SAA using (4.7). This results in a bilevel formulation. Then, using the strong duality theorem, we derive the following single-level and linear reformulation (henceforth referred to as formulation C):

$$\min \sum_{z \in Z} c_z^{\mathrm{R}} b_z p_{0z} + \sum_{z \in Z} \sum_{d \in D} c_{dz}^{\mathrm{AR}} b_z p_{dz} + \sum_{d \in D^{\mathrm{M}}} \sum_{l \in L_d} f_d^l r_d^l + \sum_{d \in D^{\mathrm{A}}} f_d x_d \tag{4.8a}$$

$$\text{s.t: } (4.4\text{b}) - (4.4\text{e}), (4.4\text{h}) - (4.4\text{k}),$$

$$\sum_{g \in D \cup \{0\}} e_{gdaks} w_{gaks} \geq x_d, \qquad d \in D, a \in A, k \in K, s \in S, \tag{4.8b}$$

$$\sum_{g \in D \cup \{0\}} e_{g0aks} w_{gaks} \geq 1, \qquad a \in A, k \in K, s \in S, \tag{4.8c}$$

$$\sum_{d \in D \cup \{0\}} w_{daks} = 1, \qquad a \in A, k \in K, s \in S, \tag{4.8d}$$

$$w_{daks} \leq x_d, \qquad d \in D, a \in A, k \in K, s \in S, \tag{4.8e}$$

$$p_{dz} \leq \frac{1}{|S|} \sum_{s \in S} \sum_{a \in A_z} \sum_{k \in K} \frac{b_{zak}}{b_z} w_{daks}, \qquad d \in D, z \in Z, \tag{4.8f}$$

$$p_{0z} \geq \frac{1}{|S|} \sum_{s \in S} \sum_{a \in A_z} \sum_{k \in K} \frac{b_{zak}}{b_z} w_{0aks}, \qquad z \in Z, \tag{4.8g}$$

$$0 \leq w_{daks} \leq 1, \qquad d \in D \cup \{0\}, a \in A, k \in K, s \in S. \tag{4.8h}$$

The objective function (4.8a), sets of constraints as well as restrictions on the decision variables are the same as in (4.4), except that Constraints (4.8b)- (4.8h) and decision variables w_{dajs} are substituted for Constraints (4.4f).

As we show below, Constraints (4.8b) - (4.8e) result from applying the strong duality theorem to (4.7). Specifically, Constraints (4.8b) and (4.8c) enforce the assignment to the CDP having the highest utility. In such constraints, the parameter $e_{gdaks} = 1$ if $U_{gaks} \geq U_{daks}, d \in D$, and 0 otherwise. Similarly, $e_{g0aks} = 1$ if $U_{gaks} \geq U_{0aks}$, and 0 otherwise. Moreover, Constraints (4.8d) ensure that customers in subzones are assigned to a single CDP or to the home delivery service. Right-hand-sides of Constraints (4.8f) and Constraints (4.8g) act as the SAA of $\psi_{dz}(\boldsymbol{x})$ and $\psi_{0z}(\boldsymbol{x})$, respecively.

Note that Constraints (4.8b) and (4.8c) correspond to one of the most widely used CAC (see Constraints ($\mathcal{CC}$) on page 50 in Espejo et al., 2012). However, in our case, we seek utility-maximizing assignments as opposed to the classic distance-minimizing assignments.

In the following, we show that the CAC-constraints given in (4.8b), (4.8c) with Constraints (4.8d) and (4.8e), are derived from the follower problem (4.7). To this end, we introduce the dual formulations of (4.6) and (4.7). They are given by

$$\min \alpha_{aks} \tag{4.9a}$$

$$\text{s.t. } \alpha_{aks} \geq U_{daks} x_d, \qquad d \in D, \tag{4.9b}$$

$$\alpha_{aks} \geq U_{0aks}. \tag{4.9c}$$

and

$$\min \gamma_{aks} + \sum_{d \in D} \beta_{daks} x_d \tag{4.10a}$$

$$\text{s.t. } \gamma_{aks} + \beta_{daks} \geq U_{daks}, \qquad d \in D, \tag{4.10b}$$

$$\gamma_{aks} \geq U_{0aks}, \tag{4.10c}$$

$$\beta_{daks} \geq 0, \qquad d \in D. \tag{4.10d}$$

Using linear programming strong duality, we rewrite (4.6) and (4.7) as the following constraint satisfaction problems:

$$\sum_{d \in D} U_{daks} x_d w_{daks} + U_{0aks} w_{0aks} \geq \alpha_{aks} \tag{4.11}$$

$$(4.6\text{b}), (4.8\text{h}), (4.9\text{b}), (4.9\text{c})$$

and

$$\sum_{d \in D \cup \{0\}} U_{daks} w_{daks} \geq \gamma_{aks} + \sum_{d \in D} \beta_{daks} x_d \tag{4.12}$$

$$(4.7\text{b}), (4.7\text{c}), (4.8\text{h}), (4.10\text{b}), (4.10\text{c}), (4.10\text{d}),$$

respectively. We use these when proving the result in the following proposition.

Proposition 5. *For each a, k and s, Constraints in* (4.12) *are equivalent to CAC-Constraints* (4.8b)*,* (4.8c)*,* (4.8d) *and* (4.8e)*.*

Proof. First, note that Constraints (4.8d) and (4.8e) are the same as Constraints (4.7b) and (4.7c). Thus, we focus on showing that Constraints (4.12), (4.10b), (4.10c) and (4.10d) yield the CAC Constraints (4.8b) and (4.8c).

According to the strong duality theorem, Constraints (4.12) are only satisfied (at the equality) at the optimum. Hence:

$$\sum_{g \in D \cup \{0\}} U_{gaks} w_{gaks} \geq \gamma^*_{aks} + \sum_{d \in D} \beta^*_{daks} \tag{4.13}$$

$$\geq \alpha^*_{aks} \tag{4.14}$$

$$\geq \max\{U_{daks} x_d, d \in D; U_{0aks}\} \tag{4.15}$$

In inequality (4.14) we use Proposition 4 to replace the optimal objective value of (4.10) with the optimal objective value of (4.9), as from there it is evident that $\alpha^*_{aks} = \max\{U_{daks} x_d, d \in D; U_{0aks}\}$. We replace (4.13)-(4.15) with the following two sets of constraints:

$$\sum_{g \in D \cup \{0\}} U_{gaks} w_{gaks} \geq U_{daks} x_d, d \in D, \tag{4.16}$$

$$\sum_{g \in D \cup \{0\}} U_{gaks} w_{gaks} \geq U_{0aks}. \tag{4.17}$$

Note that only for $g \in D \cup \{0\}$ where $U_{gaks} \geq U_{daks}$, can w_{gaks} satisfy Constraints (4.16). Also, only those with $g \in D \cup \{0\}$ where $U_{gaks} \geq U_{0aks}$, w_{gaks} can satisfy Constraint (4.17). By using $e_{gdaks} = 1$ for $U_{gaks} \geq U_{daks}$; 0 otherwise, and $e_{g0aks} = 1$ for $U_{gaks} \geq U_{0aks}$; 0 otherwise, Constraints (4.16) and (4.17) can be written as:

$$\sum_{g \in D \cup \{0\}} e_{gdaks} w_{gaks} \geq x_d, d \in D, \tag{4.18}$$

$$\sum_{g \in D \cup \{0\}} e_{g0aks} w_{gaks} \geq 1, \tag{4.19}$$

which corresponds to Constraints (4.8b) and (4.8c). □

We close this subsection with a few remarks on the solution method. Problem (4.8) can be solved using a standard Benders decomposition method implemented in a general-purpose solver, given its suitability for large MILP problems, particularly facility location problems. We identify two strategies to partition the

variables: The *first* one is *automated partition* (AP) where integer variables are assigned to the master problem and continuous variables to the subproblems. A *second* strategy, that we call *user-partition* (UP), includes also the continuous variables p_{dz} in the master problem. By doing so, we can decompose the subproblems by zones. However, the resulting subproblems are feasibility subproblems because none of the variables appear in the objective function. Hence, the UP only generates feasibility cuts. We use C-AP and C-UP, to respectively refer to the approach using AP or UP to solve model (4.8).

4.6.2. CAC with Scenario Aggregation

In this section, we describe a formulation that is based on the CAC formulation of Section 4.6.1 and the SAA with *aggregation* model (SAAA) introduced in Legault and Frejinger (2023). In this work, they achieve strong computational performance by aggregating scenarios that lead to the same solution. Moreover, they propose a partial Benders decomposition where they only explicitly account for scenarios that contribute sufficiently to the objective function. Their solution approach, however, cannot be directly applied to our model because our objective function is not submodular, and we have capacity constraints. In the following, we explain how we can nevertheless define scenario aggregation in a way that is similar to theirs.

Let (a,k,s) for $a \in A_z$, $k \in K$ and $s \in S$ denote a triplet. For each (a,k,s) we have a vector of utilities $\boldsymbol{U_{aks}} = [U_{1aks}, U_{2aks}, \ldots, U_{Daks}, U_{0aks}]$. For each vector $\boldsymbol{U_{aks}}$, we associate a vector Π_{aks} denoting the ranking of the corresponding utilities. For example, for $|D| = 2$, let us assume the following subset of triplets (a,k,s): $\{(1,1,1),(1,1,2),(1,2,1),(3,3,2)\}$, associated with the corresponding utilities: $\{[4,7,3],[2,5,1],[3,6,1],[2,5,0]\}$. Accordingly, $\Pi_{111} = \Pi_{112} = \Pi_{121} = \Pi_{332} = [2,3,1]$. Consider for instance the utility vector [4,7,3], in this case the utilities of 4, 7 and 3, have rankings 2, 3 and 1, respectively. All of the utility vectors above have the same ranking.

For the purposes of aggregation, we define a pattern q that identifies a subset of triplets (a,k,s) such that the associated vector of ranked utilities, Π_{aks}, are equal. Let Q denote the set of unique patterns. Moreover, let ν_{qak} denote the fraction of

scenarios $s \in S$ in which customers in subzone $a \in A_z$ in category $k \in K$ appear in pattern $q \in Q$. Compared to the previous formulation, we slightly change the definition of our decision variables w_{dq} that now equal 1 if the observed pattern q is assigned to CDP d, and 0 otherwise. In addition, we use $e_{gdq} = 1$, if $U_{gq} \geq U_{dq}$ and 0, otherwise, for g,$d \in D \cup \{0\}$. This leads to a formulation with aggregation (referred to as AC) where the difference with respect to (4.8) resides in the use of patterns $q \in Q$ instead of scenarios $s \in S$:

$$\min \sum_{z \in Z} c_z^{\mathrm{R}} b_z p_{0z} + \sum_{z \in Z} \sum_{d \in D} c_{dz}^{\mathrm{AR}} b_z p_{dz} + \sum_{d \in D^{\mathrm{M}}} \sum_{l \in L_d} f_d^l r_d^l + \sum_{d \in D^{\mathrm{A}}} f_d x_d \tag{4.20a}$$

s.t: (4.4b) – (4.4e), (4.4h) – (4.4k),

$$\sum_{g \in D \cup \{0\}} e_{gdq} w_{gq} \geq x_d, \qquad \forall d \in D, q \in Q, \tag{4.20b}$$

$$\sum_{g \in D \cup \{0\}} e_{g0q} w_{gq} \geq 1, \qquad \forall q \in Q, \tag{4.20c}$$

$$\sum_{d \in D \cup \{0\}} w_{dq} = 1, \qquad \forall q \in Q, \tag{4.20d}$$

$$w_{dq} \leq x_d, \qquad \forall d \in D, q \in Q, \tag{4.20e}$$

$$p_{dz} \leq \sum_{q \in Q} \sum_{a \in A_z} \sum_{k \in K} \frac{b_{zak}}{b_z} \nu_{qak} w_{dq}, \qquad \forall d \in D, z \in Z, \tag{4.20f}$$

$$p_{0z} \geq \sum_{s \in S} \sum_{a \in A_z} \sum_{k \in K} \frac{b_{zak}}{b_z} \nu_{qak} w_{0q}, \qquad z \in Z, \tag{4.20g}$$

$$0 \leq w_{dq} \leq 1, \qquad \forall d \in D \cup \{0\}, q \in Q. \tag{4.20h}$$

As with model (4.8), we solve (4.20) using a standard Benders decomposition method with AP partition (AC-AP). We note that UP is not suitable due to the linking Constraints (4.20f) and (4.20g). Indeed, the patterns $q \in Q$ may aggregate customers from different zones, hindering the decomposition over zones.

4.7. Computational Results

This section describes an extensive computational study designed to achieve two main objectives: *First*, to analyze how the levels of uncertainty in the demand model and customer locations impact the solutions. *Second*, to assess the computational performance with increasing problem size. In the following section, we describe the

experimental setup, and then discuss the results related to each of the objectives in Sections 4.7.2 and 4.7.3, respectively.

4.7.1. Experimental Setup

The experimental setup concerns the definition of the problem instances, model parameters, and performance metrics, as well as the characteristics of our infrastructure and the computing budget. We define the sets of instances by describing how we generate demand, and candidate CDP locations. For all experiments we consider a region represented by a $[0, 30] \times [0,30]$ plane, two customer categories ($|K| = 2$), and a distribution center located at [0,0] in the plane.

Demand Generation. Recall that there are two sources of uncertainty, both pertaining to demand: The LP does not have perfect knowledge about the precise location of future customers or about their preferences. To assess the impact of the first source, we generate demand with a precise location, and then we assess the impact of aggregating customers into subzones of varying sizes. As for the second source, we divide customers into two categories. Namely, 25% of the customers (drawn at random) are likelier to use CDPs than the other 75% of the customers. The strength of their preferences is governed by parameters in the utility functions as described below.

We pseudo-randomly sample $|J|$ customers' locations $(\theta_j^{\mathrm{x}}, \theta_j^{\mathrm{y}})$ within the plane ($0 \leq \theta_j^{\mathrm{x}} \leq 30$ and $0 \leq \theta_j^{\mathrm{y}} \leq 30$). Note that the superscript indicates that these correspond to x and y-coordinates in the plane. We consider two different sampling protocols. In the first (U-R, shorthand for uniform random), we assume that customers are uniformly distributed over the region. In the second (UN-R, shorthand for uniform-normal random), we assume a higher concentration of customers at the center ([15,15]) of the region. In this case, 50% of customers are uniformly distributed in the plane (same as U-R), and the other 50% are distributed according to a normal distribution θ_j^{x},$\theta_j^{\mathrm{y}} \sim N(15,3)$.

Demand Aggregation into Zones. We partition the region into equally sized zones with the same number of subzones. Since we simulate the precise location of

customers, we can aggregate all of those who are in a given zone. Hence, fewer zones implies more aggregation.

In Section 4.7.2, we use the same number of customers $|J| = 4{,}000$, and zones $|Z| = 4$, but vary the level of aggregation through the number of subzones $|A| = \{4, 8, 16, 64, 256\}$. In Section 4.7.3 where we focus on computational performance, we use $|J| = 100{,}000$ customers and vary either the number of zones – Z-class instances – or the number of subzones – A-class instances. In Z-class, we use $|Z| = \{32, 48, 64, 96\}$, and in A-class, we use $|A| = \{192, 256, 320, 400, 480, 576, 784, 1024\}$.

Note that, given the pseudo-randomly generated precise customer locations, it is possible to have instances with subzones that have no customers, or only customers of a single category. Hence, the total combination of subzones and customer categories $|\tilde{A}| \leq |A||K|$.

Next, we turn our attention to the generation of candidate locations for CDPs.

CDP Generation. We pseudo-randomly sample $|D|$ CDP locations within the plane according to a uniform distribution. For each type of CDP, the capacity is the same. We consider the 3 types of CDPs, and $|D|/3$ locations for each type, where modular parcel lockers (D^{M}) have three capacity levels ($|L| = 3$) and $L = \{u_d, 2u_d, 3u_d\}$. The minimum demand level is $\underline{b}_d = 0.5u_d$ and fixed cost $f_d = 2u_d$. In Section 4.7.2, aimed at analyzing the impact of uncertainty, we fix $|D| = 15$ and $u_d = 20$, whereas we generate more challenging instances in Section 4.7.3 – referred to as D-class instances – using $|D| = \{21{,}27{,}33{,}39{,}45\}$. We use $u_d = \{550, 825, 1650\}$ for Z, A and S-classes. For the D-class, we proportionally adjust the capacities such that the total capacity is the same for all instances.

In the following we describe model parameters, starting with how we compute variable costs, followed by the RUM model.

Variable Costs. We estimate routing costs in zone $z \in Z$ with $c_z^{\mathrm{R}} = 0.1 \times \|\theta_z\|$, where $\|\cdot\|$ denotes the L_2-norm. Note that this corresponds to the Euclidean distance between the distribution center located at [0,0] and the centroid of the zone. The resulting routing cost are $c_{dz}^{\mathrm{AR}} = 0.90 \times c_z^{\mathrm{R}}$, for $d \in D^{\mathrm{P}}$, and $c_{dz}^{\mathrm{AR}} = 0.80 \times c_z^{\mathrm{R}}$, for $d \in D \setminus D^{\mathrm{P}}$.

RUM Model. We postulate a logit model in all experiments, i.e., ε_{daks} are independently and identically distributed according to the Extreme Value type I distribution with scale parameter β. We use a random utility defined as $U_{daks} = \alpha_k^1\big(|\theta_a^{\mathrm{x}} - \theta_d^{\mathrm{x}}| + |\theta_a^{\mathrm{y}} - \theta_d^{\mathrm{y}}|\big) + \varepsilon_{akds}$. Here, again, the superscripts denote the x- and y-coordinates in the plane. We use $\alpha_2^1 = 5\alpha_1^1$ so that customers in category $k = 2$ are five times more sensitive to the distance than customers in category $k = 1$. Hence, the values of β and α_1^1 determine the level of uncertainty and customers' preferences.

In Section 4.7.2, we use three configurations of these parameters $(\alpha_1^1, \beta) = \{(-0.01, 1), (-0.1{,}1), (-0.1, 0.25)\}$, whereas in Section 4.7.3 we use a subset of two configurations $(\alpha_1^1, \beta) = \{(-0.1, 1), (-0.1, 0.25)\}$.

Instances and Scenarios. Our formulations are based on scenarios obtained by sampling values of the random terms. For each of the parameter configurations that we have described, we draw ten sets of scenarios. An instance refers to a problem defined by a set of scenarios, hence we have ten instances for each parameter setting.

Moreover, we consider different numbers of scenarios. In Section 4.7.2, we use $|S| = 50$, and we also keep it fixed to this number when solving the Z-, A- and D-classes of instances. To assess the impact on computing time, we consider a fourth class of instances, referred to as the S-class where we vary the number of scenarios $|S| = \{50, 100, 150, 200, 250, 300\}$.

Performance Metrics and Interpretability. We report different metrics depending on the focus of the analysis, including computing time in seconds, the number of CDPs that are open in the optimal solution, the value of the objective function, the optimality gap for instances that are not solved to optimality within our time limit. For interpretability of the level of uncertainty associated with the demand model, we use the entropy measure introduced in Legault and Frejinger (2023). The entropy is defined as $H = \sum_{q \in Q} -v_q \log(v_q)$, where v_q is the fraction of scenarios where pattern q is observed. We report entropy, or average entropy $\hat{H}$. Note that the maximum entropy (i.e. the entropy of a uniform distribution) may vary across instances. Therefore, we also report the maximum entropy and categorize instances as having low, medium or high entropy relative to the maximum.

Infrastructure and Computing Budget. The experiments are performed on the Linux version of IBM ILOG CPLEX 22.1, running on an Intel I7-9700K 8 cores CPU at 3.6GHz/4.9GHz, with a time budget of one hour.

4.7.2. Impact of Uncertainty

In this section, we analyze two aspects that impact the level of uncertainty. Namely, (i) parameters of the RUM model, and (ii) geographical distribution of CDPs and customer aggregation into subzones. Recall that we keep the numbers of customers, zones and CDP fixed for the results in this section, while we use different configurations of the utility functions and different numbers of subzones. Figures 4.2 – 4.4 display different views on the same set of instances.

Figure 4.2 is a scatter plot showing, for each instance, the number of open CDPs in the optimal solution and the associated entropy level expressed as a ratio of the maximum entropy. All 15 CDPs are open in most high-entropy instances. High entropy implies that all CDPs have similar estimated probabilities to capture demand. In this case, the total reduction in estimated routing costs surpasses the fixed cost. Moreover, all CDPs have the same fixed cost. Therefore, the model chooses to open all CDPs.

The figure clearly shows three clusters of instances: low, medium, and high entropy. Note that we include high-entropy configurations in these results for the sake of illustration. In practice, such configurations are of less interest as they imply demand models that do not provide any information compared to a uniform distribution. In the following we provide a more in-depth analysis of the low and medium entropy instances.

Figures 4.3a and 4.3b display the distributions of the number of open CDPs in the optimal solutions as a function of parameter configurations. The latter is the geographical distribution of demand (uniform, U-R, and more concentrated to the center UN-R), and the number of subzones $|A|$. We note that all 15 CDPs are open in most of the medium entropy instances except those having the fewest number of subzones. We see a higher variance in the low-entropy instances and fewer open CDPs. The median (indicated with a line in each box) is higher for larger number

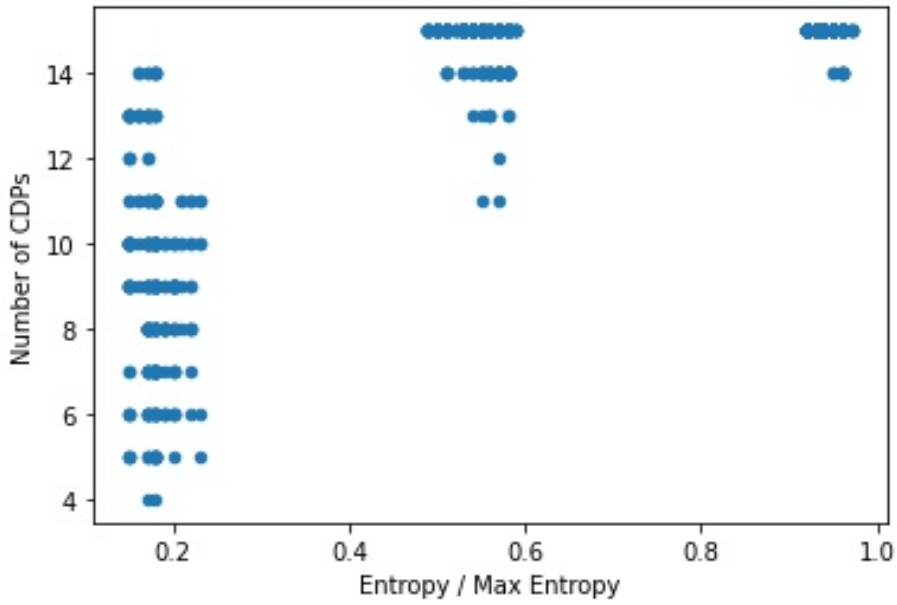


Fig. 4.2. Number of open CDPs in the optimal solution for different values of the ratio between entropy and max entropy for each instance

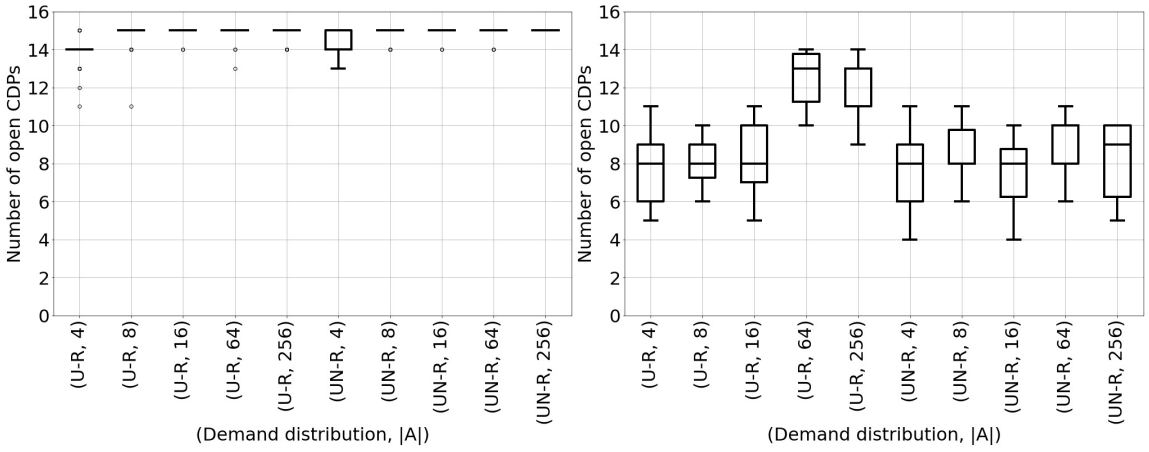
of subzones, especially when demand is uniformly distributed in the region. Now, turning to the optimal expected costs displayed similarly in Figures 4.4a and 4.4b, we note that they decrease as the number of subzones increases. Moreover, they are smaller when demand is more highly concentrated in the center of the region (UN-R compared to U-R). In the following, we provide an analysis of these results.

Recall that the model trades off reduction in variable routing costs with fixed costs associated with opening CDPs. The reduction of variable routing costs depends on the demand that a CDP can attract which, in turn, is governed by the utility functions. In the low-entropy instances, attraction strongly depends on distance, whereas this dependence is less strong in medium-entropy instances.

We use the centroid of a subzone to compute distances. The distribution of demand in the region, along with the size of the subzones, therefore determine the quality of the distance approximation. Consider uniformly distributed demand in the region. Having only four subzones means that distance will be overestimated for a significant share of the demand, compared to having smaller (more) subzones.

This issue is less pronounced when demand is concentrated to the center of the region (UN-R distribution).

Given the above, less demand is attracted by CDPs, especially in low-entropy instances when subzones are relatively large (see Figure B.1 in Appendix B.3). This, hence, explains why those instances have fewer open CDPs and higher costs. In other words, it is valuable to reduce uncertainty by trying to model demand as accurately as possible (utility functions and geographical distribution), and to solve the model with a larger number of subzones. Of course, from a practical point of view, the size of the subzones should reflect the level of uncertainty in the geographical distribution of demand. It does not make sense to have small subzones if there is a high degree of uncertainty in where the demand is located. As expected, having areas with relatively high concentration of demand leads to lower costs. Next we turn our attention to an analysis of computing time for larger instances.

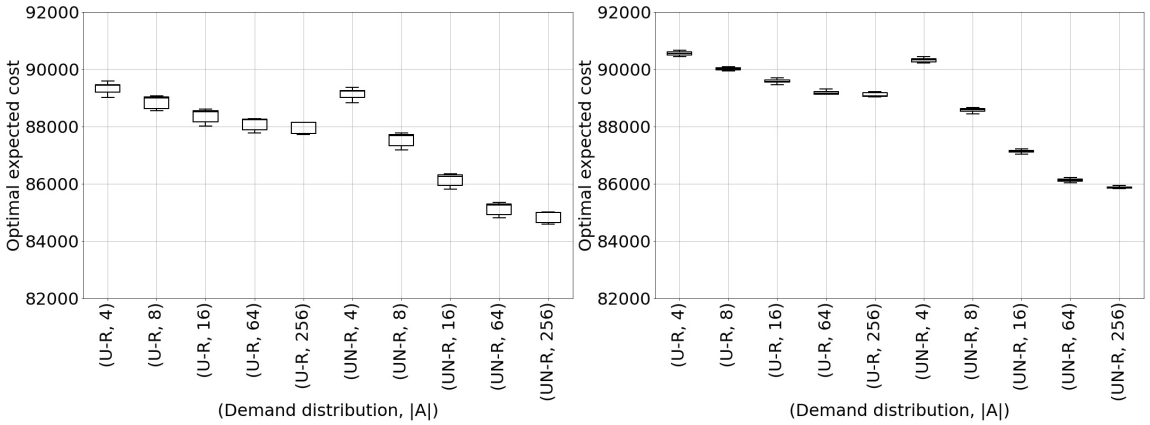


(a) Medium entropy **(b)** Low entropy

Fig. 4.3. Distribution of the optimal number of CDPs

4.7.3. Model and Algorithm Comparison

In this section we analyze the computing time when the size of the instances increases in various dimensions. We compare the three approaches C-AP, C-UP and



(a) Medium level entropy **(b)** Low level entropy

Fig. 4.4. Distribution of the optimal expected cost

AC-AP. As detailed in Section 4.7.1, we use four classes of instances for this purpose: S-, Z-, A- and D-classes.

Tables 4.4 – 4.7 report the average computing time (in seconds) for the four classes of instances. Recall that the average in each row is computed over ten instances. All instances are solved to optimality with only 10 exceptions (three configurations of the D-class, see footnotes in Table 4.7). In addition to computing time, we report average entropy ($\hat{H}$), and average maximum entropy (avg max H). Each table is separated into two parts, the upper reports results for medium-entropy instances, and the lower for low-entropy instances. The shortest computing time in each row is highlighted in bold font.

A few findings clearly emerge from these results. The AC-AP is the best approach for low-entropy instances across all classes. This is expected, and consistent with the findings in Legault and Frejinger (2023), as aggregation has the largest impact on model size for low-entropy instances. We note that AC-AP dominates the other approaches for each individual instance and not only on average (see Figure B.2 in Appendix B.3). In terms of absolute computing time, low-entropy instances are also the easiest to solve. This opens up the possibility to use large number of subzones and scenarios for improved modeling accuracy.

Next, we focus on findings related to medium-entropy instances. First, we note that increasing the size of Z marginally impacts the computing time as this dimension only affects continuous variables p_{dz}. The hardest instances arise when increasing $|D|$ as it governs the number of binary variables x_d. Instances where the capacity is not restrictive (larger u_d) are especially challenging. When capacities are restrictive, more demand is allocated to home delivery service, which makes the problem easier to solve. Computing time also increases with $|S|$ and $|A|$ but less strongly than increasing $|D|$.

Unlike for low-entropy instances, a single approach does not dominate the others. For the S-class, C-UP performs the best as $|S|$ increases. There is no clear winner for the A-class instances but the computing times are quite similar for the three approaches, except for the two largest configurations where AC-AP is significantly faster on average. On the hardest class of instances (D-class), C-AP performs the best overall. C-UP has a lower average computing time on the largest instances but solves only one out of ten instances to optimality within the time limit. For the two configurations where AC-AP has the lowest average computing time, C-AP has a similar performance.

The entropy of an instance can be computed before it is solved. The results indicate that the solution approach should be selected based on this information. For low-entropy instances, the best approach is to use aggregation AC-AP. For medium (to high) entropy instances, C-AP consistently performs well, if not the best. Especially for the hardest instances with relatively high number of candidate CDP locations. However, if the objective is to improve modeling accuracy by increasing the number of scenarios and the number of subzones, then C-UP can achieve a strong performance. We close this section by noting that the approaches we propose based on reformulations using (4.8) consistently display shorter computing times (sometimes by several orders of magnitude) compared to those that result from using (4.6). We provide more details in Appendices B.2 and B.3.

				Avg. computing time (sec)		
$\hat{H}$	avg max H	$\|S\|$	u_d	**C-AP**	**C-UP**	**AC-AP**
4.5	8.8	50	550	**2.0**	2.3	2.2
4.5	8.8	50	825	**2.4**	2.6	2.6
4.5	8.8	50	1650	**3.0**	4.4	3.3
4.8	9.4	100	550	**5.4**	5.8	6.7
4.8	9.4	100	825	**6.4**	6.7	7.5
4.8	9.4	100	1650	**8.7**	9.5	10.1
5.0	9.9	150	550	**10.6**	**10.6**	13.8
5.0	9.9	150	825	12.7	**11.9**	16.1
5.0	9.9	150	1650	**17.1**	17.6	19.9
5.1	10.1	200	550	16.7	**16.6**	23.4
5.1	10.1	200	825	19.8	**18.7**	26.0
5.1	10.1	200	1650	27.8	**27.6**	33.3
5.2	10.4	250	550	25.3	**24.1**	35.7
5.2	10.4	250	825	30.0	**26.8**	39.9
5.2	10.4	250	1650	41.0	**40.2**	50.2
5.3	10.5	300	550	35.1	**33.9**	49.5
5.3	10.5	300	825	41.2	**38.3**	55.5
5.3	10.5	300	1650	57.5	**53.8**	71.2
5.4	10.7	350	550	47.2	**44.2**	69.0
5.4	10.7	350	825	54.3	**48.5**	76.3
5.4	10.7	350	1650	78.5	**68.2**	101.4
5.4	10.8	400	550	59.8	**55.7**	98.5
5.4	10.8	400	825	70.0	**61.5**	108.8
5.4	10.8	400	1650	105.3	**87.2**	142.3
1.2	7.4	50	550	0.5	0.6	**0.1**
1.2	7.4	50	825	0.5	0.5	**0.1**
1.2	7.4	50	1650	0.5	0.5	**0.1**
1.2	8.0	100	550	1.2	1.2	**0.4**
1.2	8.0	100	825	1.1	1.1	**0.3**
1.2	8.0	100	1650	1.1	1.1	**0.3**
1.3	8.5	150	550	1.9	1.9	**0.6**
1.3	8.5	150	825	1.8	1.8	**0.5**
1.3	8.5	150	1650	1.7	1.7	**0.5**
1.3	8.7	200	550	2.7	2.6	**0.9**
1.3	8.7	200	825	2.5	2.5	**0.8**
1.3	8.7	200	1650	2.4	2.4	**0.7**
1.3	9.0	250	550	3.6	3.6	**1.2**
1.3	9.0	250	825	3.4	3.4	**1.1**
1.3	9.0	250	1650	3.2	3.2	**1.0**
1.4	9.1	300	550	4.7	4.7	**1.6**
1.4	9.1	300	825	4.5	4.5	**1.5**
1.4	9.1	300	1650	4.1	4.1	**1.3**
1.4	9.3	350	550	5.9	5.8	**2.0**
1.4	9.3	350	825	5.6	5.6	**1.8**
1.4	9.3	350	1650	5.0	5.0	**1.6**
1.4	9.4	400	550	7.3	7.2	**2.5**
1.4	9.4	400	825	6.8	6.8	**2.2**
1.4	9.4	400	1650	6.2	6.3	**2.0**

Table 4.4. Average computing time for S-class instances.

4.8. Case Study

This section presents a case study where the LP seeks to implement CDPs in the region of Montreal, Quebec. For this study, we utilize data from our industrial

				Avg. computing time (sec)		
$\hat{H}$	avg max H	$\lvert\tilde{A}\rvert$	u_d	**C-AP**	**C-UP**	**AC-AP**
4.7	9.0	384	550	**2.9**	3.3	3.5
4.7	9.0	384	825	**3.7**	4.0	4.2
4.7	9.0	384	1650	**4.8**	6.5	5.3
4.8	9.3	512	550	**4.6**	4.8	5.7
4.8	9.3	512	825	**5.6**	**5.6**	6.7
4.8	9.3	512	1650	**7.3**	8.7	8.1
4.9	9.6	640	550	**6.2**	6.6	8.1
4.9	9.6	640	825	**7.6**	7.7	9.2
4.9	9.6	640	1650	**10.4**	11.5	11.9
5.0	9.8	800	550	**9.4**	**9.4**	12.2
5.0	9.8	800	825	11.0	**10.7**	14.0
5.0	9.8	800	1650	**15.1**	15.8	18.1
5.1	10.0	959	550	12.2	**12.1**	16.6
5.1	10.0	959	825	14.3	**13.8**	18.4
5.1	10.0	959	1650	**20.7**	21.4	24.8
5.1	10.1	1151	550	**15.8**	16.2	22.6
5.1	10.1	1151	825	19.3	**18.7**	25.7
5.1	10.1	1151	1650	28.4	**27.1**	35.1
5.3	10.4	1553	550	27.4	**26.9**	40.4
5.3	10.4	1553	825	**30.1**	**30.1**	44.1
5.3	10.4	1553	1650	58.0	**54.5**	60.1
5.5	10.6	1800	550	32.4	**32.2**	56.9
5.5	10.6	1800	825	364.0	103.3	**65.0**
5.5	10.6	1800	1650	255.9	226.5	**88.7**
1.2	7.6	384	550	0.8	0.8	**0.2**
1.2	7.6	384	825	0.7	0.7	**0.2**
1.2	7.6	384	1650	0.7	0.7	**0.2**
1.2	7.9	512	550	1.0	1.1	**0.3**
1.2	7.9	512	825	1.0	1.0	**0.3**
1.2	7.9	512	1650	1.0	1.0	**0.2**
1.3	8.2	640	550	1.3	1.3	**0.4**
1.3	8.2	640	825	1.2	1.3	**0.3**
1.3	8.2	640	1650	1.2	1.2	**0.3**
1.3	8.4	800	550	1.6	1.7	**0.5**
1.3	8.4	800	825	1.6	1.6	**0.5**
1.3	8.4	800	1650	1.6	1.6	**0.4**
1.3	8.6	959	550	2.0	2.0	**0.7**
1.3	8.6	959	825	1.9	1.9	**0.6**
1.3	8.6	959	1650	1.9	1.9	**0.5**
1.3	8.7	1151	550	2.5	2.5	**0.9**
1.3	8.7	1151	825	2.3	2.4	**0.8**
1.3	8.7	1151	1650	2.3	2.3	**0.7**
1.4	9.0	1553	550	3.5	3.6	**1.4**
1.4	9.0	1553	825	3.3	3.3	**1.1**
1.4	9.0	1553	1650	3.2	3.2	**1.1**
1.4	9.2	1800	550	4.3	4.4	**1.9**
1.4	9.2	1800	825	4.0	4.0	**1.5**
1.4	9.2	1800	1650	3.9	3.9	**1.5**

Table 4.5. Average computing time for A-class instances

partner, Purolator. We transform data related to cost, demand, and preferences to preserve confidentiality. In this case, the demand is known at the postal codes, which are grouped in 35 FSA (forward sorting area). Thus, $|Z| = 35$, considering only those FSAs that are served from the main terminal. In addition, only postal codes having

				Avg. computing time (sec)		
$\hat{H}$	avg max H	$\lvert Z\rvert$	u_d	**C-AP**	**C-UP**	**AC-AP**
5.0	9.7	32	550	8.5	**8.2**	11.2
5.0	9.7	32	825	10.6	**9.6**	13.0
5.0	9.7	32	1650	**13.9**	14.3	16.6
5.0	9.7	48	550	9.1	**8.4**	11.6
5.0	9.7	48	825	10.5	**9.7**	13.2
5.0	9.7	48	1650	**14.4**	16.4	16.6
5.0	9.7	64	550	**8.9**	9.0	11.5
5.0	9.7	64	825	11.1	**9.9**	13.7
5.0	9.7	64	1650	**14.3**	17.8	16.7
5.0	9.7	96	550	**9.0**	9.8	11.6
5.0	9.7	96	825	**11.1**	11.2	13.4
5.0	9.7	96	1650	**14.1**	23.5	16.5
1.3	8.3	32	550	1.6	1.6	**0.5**
1.3	8.3	32	825	1.5	1.5	**0.4**
1.3	8.3	32	1650	1.5	1.5	**0.4**
1.3	8.3	48	550	1.6	1.7	**0.5**
1.3	8.3	48	825	1.5	1.6	**0.4**
1.3	8.3	48	1650	1.6	1.5	**0.4**
1.3	8.3	64	550	1.5	1.7	**0.5**
1.3	8.3	64	825	1.5	1.6	**0.4**
1.3	8.3	64	1650	1.5	1.5	**0.4**
1.3	8.3	96	550	1.6	1.9	**0.5**
1.3	8.3	96	825	1.6	1.7	**0.4**
1.3	8.3	96	1650	1.6	1.7	**0.4**

Table 4.6. Average computing time for Z-class instances

regular (non-zero) weekly demand are considered leading to $|A| = 1{,}081$ postal codes. The total weekly demand amounts to 29,092 parcels. We assume three categories of customers, and we make them fictitious for confidentiality reasons. Table 4.8 shows the three categories of customers and the corresponding utility parameters. The utility only depends on the distance from CDPs to the customer's corresponding location ($\alpha_k^2 = 0$). In addition, we assume that the utility function represents an accurate estimation of customer preferences with $\beta = 0.25$. Nevertheless, we illustrate the impact of using $\beta = 1$ which implies a higher degree of uncertainty. We also analyze the effect of more sensitive customers. For this case study, we impose a 10-hour computing time budget.

We consider 40 CDP candidate locations, where 10 correspond to stores (D^{P}), 15 to modular parcel lockers, and 15 to regular parcel lockers (D^{R}). In Table 4.9, we present the data related to the CDPs. Note that CDPs at stores charge an additional

				Avg. computing time (sec)		
$\hat{H}$	avg max H	$\|D\|$	u_d	**C-AP**	**C-UP**	**AC-AP**
5.4	9.5	21	393	**10.2**	10.8	13.0
5.4	9.5	21	589	**13.7**	13.9	16.7
5.4	9.5	21	1179	**14.3**	24.4	17.4
5.9	9.5	27	306	**13.1**	14.7	17.9
5.9	9.5	27	458	**16.3**	19.2	20.3
5.9	9.5	27	917	**43.6**	131.1	54.7
6.2	9.6	33	250	**15.2**	21.4	22.3
6.2	9.6	33	375	**55.5**	63.2	58.7
6.2	9.6	33	750	**288.1**	1141.1(1)	294.4
6.5	9.6	39	212	**20.8**	30.1	31.4
6.5	9.6	39	317	**46.7**	64.4	54.6
6.5	9.6	39	635	211.4	1240.4(2)	**205.0**
6.7	9.7	45	183	**27.1**	40.6	41.6
6.7	9.7	45	275	70.5	98.5	**69.3**
6.7	9.7	45	550	925.9	**669.0**(3)	900.8
1.7	8.3	21	393	1.7	1.7	**0.6**
1.7	8.3	21	589	1.6	1.7	**0.6**
1.7	8.3	21	1179	1.6	1.6	**0.5**
2.0	8.4	27	306	2.5	2.5	**1.1**
2.0	8.4	27	458	2.5	2.5	**1.0**
2.0	8.4	27	917	2.4	2.3	**0.9**
2.3	8.6	33	250	3.5	3.6	**1.6**
2.3	8.6	33	375	3.6	3.6	**1.6**
2.3	8.6	33	750	3.2	3.2	**1.3**
2.6	8.7	39	212	4.5	4.5	**2.2**
2.6	8.7	39	317	4.7	4.8	**2.3**
2.6	8.7	39	635	4.4	4.4	**1.9**
2.8	8.8	45	183	5.7	5.7	**2.8**
2.8	8.8	45	275	5.8	5.9	**2.9**
2.8	8.8	45	550	5.9	6.0	**2.7**

(1) 9/10 solved to optimality, 0.04% gap for 1/10
(2) 9/10 solved to optimality, 0.02% gap for 1/10
(3) 1/10 solved to optimality, 0.06% average gap for 9/10

Table 4.7. Average computing time for D-class instances

Customer category	**% Population**	α_k^1
Category 1	20%	-0.10
Category 2	70%	-0.15
Category 3	10%	-0.25

Table 4.8. Customer categories and utility function parameters

\$1 for delivering parcels to customers. In addition, f_d correspond to weekly amortized fixed costs, and u_d corresponds to the capacity of CDPs for a week. Figure 4.5 shows the centroid location of postal codes and the potential CDPs for this case.

Without considering CDPs, the (current) estimated total routing cost per week is \$73,301. This cost is reduced by 3.1% in the optimal solution. In this case, 7 regular parcel lockers are open, capturing 26.4% of the demand. Noteworthy is that even if

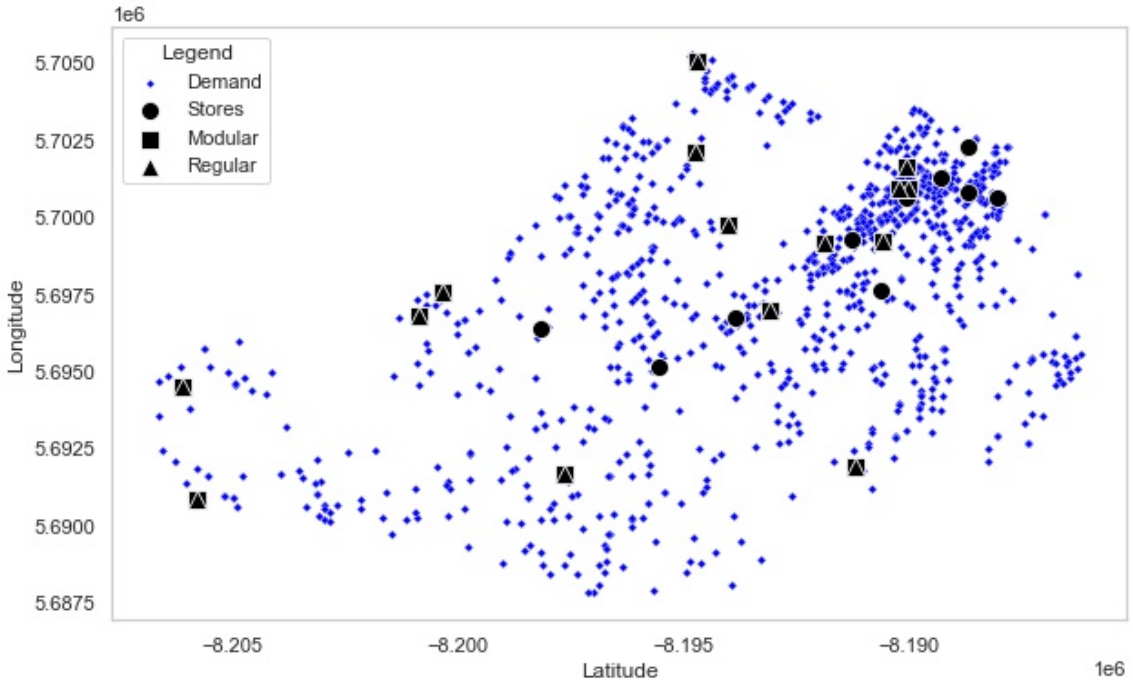


Fig. 4.5. Postal Codes and CDPs.

modular parcel lockers display a better unit cost capacity (f_d/u_d) than regular parcel lockers, they are not selected. However, the much lower capacity of modular parcel lockers limits the total cost reductions they can provide.

If, instead, customers in the first and second categories are 25% more sensitive to distance, only 6 regular parcel lockers are needed, reducing the current estimated total costs by 2.3% and capturing 21.9% of the demand. On the contrary, a choice model with $\beta = 1$ significantly overestimates the demand captured by CDPs (56%) which leads to opening 14 CDPs.

We note that regular parcel lockers are preferred over the other two classes of CDPs, but this is mainly due to their associated costs and capacity. If the cost for delivering a parcel by a store were \$0.5 instead of \$1, then 9 CDPs are open at stores, and 5 regular parcel lockers are implemented, resulting in a routing cost reduction of 3.7%. In addition, in such a case, the 14 CDPs capture 36.1% of the demand.

In summary, CDPs' capacities not only limit the demand they can attract but also negatively impact the potential cost reductions they offer. In addition, underestimating customers' negative perception of distance (leading to higher uncertainty) may lead to an overestimation of the number of CDPs.

Type of CDP	u_d	f_d	b_d	$c_{dz}^{\mathbf{AR}}$
Stores	600.0	0.0	300.0	$0.7 \times c_z^{\mathrm{R}} + 1$
Modular	148	31.0	0.0	$0.7 \times c_z^{\mathrm{R}}$
Regular	1200	500.0	0.0	$0.7 \times c_z^{\mathrm{R}}$

Table 4.9. CDP instance data

4.9. Conclusions

We addressed a choice-based capacitated facility location problem with random utility-maximizing customers. We considered the case of an LP that seeks to locate CDPs of different types. At this strategic level, the LP has imperfect knowledge of customer preferences as well as precise customer locations. The CDP locations and, hence, the benefit for the LP of these facilities depend on the capacity to attract customers. Thus, customer locations and preferences for the CDPs are key drivers for the decisions.

We proposed two formulations including closest assignment constraints that we solved using a standard Benders decomposition method. We showed that the closest assignment constraints are derived from a single-level reformulation of a bilevel program with a specific formulation for the follower's problem. We conducted an extensive experimental study to analyze the impact of the level of uncertainty (customer preferences and geographical distribution) on the optimal solutions and computing time. We used an entropy measure for interpretability. Consistent with the finding in Legault and Frejinger (2023), the most effective model and solution approach depends on the level of uncertainty in the instance. This is captured by the entropy and, importantly, it can be easily computed before solving an instance. Using scenario aggregation is the best option for low-entropy instances. For higher entropy levels, the formulation without aggregation resulted in shorter computing times. In

this case, we considered two ways to define the subproblems: the standard (referred to as automatic partition) way where integer variables are in the master and continuous variables in the subproblems, and user partition where we also include continous variables in the master resulting in exclusively generating feasibility cuts. The user partition worked particularly well when increasing the number of scenarios, whereas the automatic partition displayed strong performance more consistently across all instances.

The experimental results showed that the results are quite sensitive to the demand model and accuracy of the geographical distribution of demand. This highlights the need for high-performing demand models, especially in regions where it impacts the decisions (and objective function) the most. We believe that integrated learning and optimization (Sadana et al., 2025) for training such models constitute a potentially valuable direction for future research.

In this work we assumed that demand is allocated to home delivery if there is insufficient capacity at the preferred CDP. Relaxing this assumption would lead to a more complex setting which we intend to explore in future research.

Conclusion

This thesis has examined the challenges encountered by an expected profit-maximizing or expected cost-minimizing LP in its strategic decisions about capacity planning, pricing, and facility location, in a context where only incomplete information is available about the demand it is facing.

Our methodological contributions are the following: In the *first* article (Chapter 2), we expressed the LP's expected profit-maximizing problem through the combination of an optimization framework with a random choice model and we transformed its initial formulation as a stochastic bi-level program into a deterministic single-level program, by using duality and SAA. Then, we proposed a heuristic approach H-A based on a reduced single-level program that performed particularly well in contexts where the uncertainty resulting from the choice model configurations is moderate. This method significantly outperformed the BA approach that consists in solving the SAA formulation with Benders decomposition. In the *second* article (Chapter 4), we proposed two simulation-based asymptotically exact methods to solve a capacitated facility location problem involving random utility-maximizing customers. These methods were based on a formulation resulting from the introduction of closest assignment constraints and scenario aggregation and performed Benders decomposition under two alternative partitions. We specified the conditions of application of each method. In the *third* article (Chapter 3), we improved on the BA method developed in the first paper by applying a scenario aggregation procedure inspired from Legault and Frejinger (2023). The resulting pattern-based reformulation yielded the asymptotically exact BA-R method whose speed exceeded that of

the reference BA by several orders of magnitude. This made it possible to extend the set of instances that could be solved to optimality and therefore to evaluate more extensively the heuristic methods proposed in the second article.

We also provided managerial insights into the effects of the random utility choice model on optimal pricing, facility location, and capacity decisions by evaluating the performance of our computational methods in large-scale, realistic experiments. In the *first* paper, we showed that the optimal expected profit was sensitive to the parameters of the choice model, namely to the sensitivity of preferences to prices and to the level of uncertainty surrounding the estimates of utility. In the *second* paper, we showed that the optimal number of CDPs and the achievable reduction in expected costs are sensitive to the level of uncertainty about the customers' preferences and geographical distribution.

Our results are subject to a number of limitations: *First*, the formulations and the performances of the proposed approaches are tied to the RUM class of demand models. However, limited information is available to guide the selection of particular model instances and model parameters. Therefore, their determination must be based partly on reasonable postulates and partly on statistical estimates. This is the case about the preferences over the combinations of price and service level. *Second*, the exact approaches - BA and BA-R - are affected by symmetry issues impinging on computational performance. *Third*, as the problems that we addressed are relatively new, we focused on their essential characteristics and relaxed some business constraints to yield simpler formulations. The applicability of the results in more complex contexts of application should be verified. As a first example, we assumed in the first paper that the LP independently addresses each shipper's demand to offer specific price and service level. However, the demands of several shippers might be addressed jointly. As a second example, we assumed in the second paper that the demand is necessarily allocated to home delivery if there is insufficient capacity at the preferred CDP. Yet, the next preferred CDP might potentially accommodate the demand in excess or in its totality. *Fourth*, for instances of very large size, handling the proposed formulations, especially those of the second paper, requires a commensurately large amount of computing memory.

In view of the stated findings and limitations, we envision the following directions for future research:

Demand Model. As a response to the dependency of the optimal solutions on the characteristics of the RUM models, systematic research about the selection of particular instances of the RUM class and about the estimation of the model parameters would be desirable. Integrating learning (i.e., estimation) and optimization (see Sadana et al., 2025) would also be worthy of exploring.

Extensions. The formulation of the first article could be extended so that the individual demands faced by the LP can be aggregated in sets whose members' demands are jointly rather than independently addressed. Also, service levels could be more finely defined. Thus, given a set of attributes and the customers' preferences towards them, a decision would be made on the optimal level of each attribute in a service level offered to a shipper. In the second article, the formulation could be extended so that excess demand or total demand might be diverted to the next preferred CDP once the capacity of the CDP with the next higher utility is surpassed.

Computational Performance. We distinguish two possible subareas of research in this regard: *First*, an improvement in computational performance can be sought through *model enhancements*. Since the introduction of symmetry-breaking inequalities would boost the performance of the methods presented in the first and third articles, we find it worthwhile to investigate the construction of valid inequalities based on the information available about simulated utilities. *Second*, an improvement in computational performance can also be sought through *algorithmic enhancements*. (i) Improvements to decomposition methods (to Benders decomposition in particular) would make it possible to handle larger problems and to reduce memory requirements in a context where applications steadily grow in size. (ii) Similarly, column generation with simulated utilities could improve the tractability of very large problems and reduce memory requirements. In this case, each column could be tied to an agreement profile or a subset of open CDPs and therefore to the corresponding simulated choice probability. A main challenge consists in selecting the most promising columns to accelerate the optimization process. (iii) Also, we can further improve

the BA-R method by defining an approach that avoids the repetitive evaluation of agreement patterns leading to the same embedded CFLP.

Choice-based optimization emerges as a promising modeling approach for several strategic/tactical decisions in logistics. Furthermore, logistic operations, particularly last-mile operations, require the use of more disaggregated demand levels. Hence, choice-based optimization approaches can bring about valuable insights, but not without an increase in computational resources, and in even greater amount if the problems addressed involve two or more decision-makers with conflicting objectives. This constitutes additional motivation for research on algorithmic and modeling enhancements.

References

Abbasi, S., Çiğdem Sıcakyüz, Santibanez Gonzalez, E. D., and Ghasemi, P. A systematic literature review of logistics services outsourcing. *Heliyon*, 10(13):e33374, 2024. doi: https://doi.org/10.1016/j.heliyon.2024.e33374.

Alizadeh, S. M., Marcotte, P., and Savard, G. Two-stage stochastic bilevel programming over a transportation network. *Transportation Research Part B: Methodological*, 58:92 – 105, 2013. doi: https://doi.org/10.1016/j.trb.2013.10.002.

Amorim, P., DeHoratius, N., Eng-Larsson, F., and Martins, S. Customer preferences for delivery service attributes in attended home delivery. *Management Science*, 0 (0), 2024. doi: 10.1287/mnsc.2020.01274.

Arslan, O. The location-or-routing problem. *Transportation Research Part B: Methodological*, 147:1–21, 2021. doi: https://doi.org/10.1016/j.trb.2021.02.010.

Bard, J. F. *Practical Bilevel Optimization: Algorithms and Applications*. Springer, 1998.

Bazotte, M., Carvalho, M., and Vidal, T. Solving two-stage stochastic programs with endogenous uncertainty via random variable transformation, 2024. URL `https://arxiv.org/abs/2402.15486`.

Beasley, J. E. Or-library: Distributing test problems by electronic mail. *The Journal of the Operational Research Society*, 41(11):1069–1072, 1990. doi: https://doi.org/10.2307/2582903.

Beasley, J. An algorithm for solving large capacitated warehouse location problems. *European Journal of Operational Research*, 33(3):314–325, 1988. doi: https://doi.org/10.1016/0377-2217(88)90175-0.

Beck, Y., Ljubić, I., and Schmidt, M. A survey on bilevel optimization under uncertainty. *European Journal of Operational Research*, 311(2):401–426, 2023. doi: https://doi.org/10.1016/j.ejor.2023.01.008.

Ben-Akiva, M. and Lerman, S. R. *Discrete Choice Analysis: Theory and Application to Travel Demand*. MIT Press, Cambridge, MA, 1985.

Benati, S. and Hansen, P. The maximum capture problem with random utilities: Problem formulation and algorithms. *European Journal of Operational Research*, 143(3):518–530, 2002. doi: 10.1016/S0377-2217(01)00340-X.

Beresnev, V. L. and Melnikov, A. A. A capacitated competitive facility location problem. *Journal of Applied and Industrial Mathematics*, 10(1):61–68, 2016. doi: 10.1134/S1990478916010075.

Beresnev, V. and Melnikov, A. Exact method for the capacitated competitive facility location problem. *Computers & Operations Research*, 95:73–82, 2018. doi: 10.1016/j.cor.2018.02.013.

Berman, O., Huang, R., Kim, S., and Menezes, M. B. C. Locating capacitated facilities to maximize captured demand. *IIE Transactions*, 39(11):1015–1029, 2007. doi: 10.1080/07408170601142650.

Bertsimas, D. and Mundru, N. Optimization-based scenario reduction for data-driven two-stage stochastic optimization. *Operations Research*, 71(4):1343–1361, 2023. doi: 10.1287/opre.2022.2265.

Birge, J. R. and Louveaux, F. *Introduction to stochastic programming*. Springer series in operations research and financial engineering. Springer, second edition. edition, 2011.

Brotcorne, L., Labbé, M., Marcotte, P., and Savard, G. A bilevel model and solution algorithm for a freight tariff-setting problem. *Transportation Science*, 34:289–302, 2000. doi: https://doi.org/10.1287/trsc.34.3.289.12299.

Buchheim, C. Bilevel linear optimization belongs to np and admits polynomial-size kkt-based reformulations. *Operations Research Letters*, 51(6):618–622, 2023. doi: https://doi.org/10.1016/j.orl.2023.10.006.

Camacho-Vallejo, J.-F., Corpus, C., and Villegas, J. G. Metaheuristics for bilevel optimization: A comprehensive review. *Computers & Operations Research*, 161:

106410, 2024. doi: https://doi.org/10.1016/j.cor.2023.106410.

Casas-Ramirez, M.-S., Camacho-Vallejo, J.-F., and Martinez-Salazar, I.-A. Approximating solutions to a bilevel capacitated facility location problem with customer's patronization toward a list of preferences. *Applied Mathematics and Computation*, 319:369 – 386, 2018. doi: https://doi.org/10.1016/j.amc.2017.03.051.

Chen, L. Identifying effective scenarios for sample average approximation, 2019. URL `https://arxiv.org/abs/1904.01550`.

Choo, C. *Impact of a delivery point network for urban e-commerce deliveries*. PhD thesis, 2016.

Chopra, S. and Meindl, P. *Supply Chain Management: Strategy, Planning, and Operation*. Pearson Education, 6th edition, 2016.

Colson, B., Marcotte, P., and Savard, G. An overview of bilevel optimization. *Annals of Operations Research*, 153(1):235–256, 2007. doi: 10.1007/s10479-007-0176-2.

Cornuejols, G., Sridharan, R., and Thizy, J. A comparison of heuristics and relaxations for the capacitated plant location problem. *European Journal of Operational Research*, 50(3):280–297, 1991. doi: https://doi.org/10.1016/0377-2217(91)90261-S.

Dan, T., Lodi, A., and Marcotte, P. Joint location and pricing within a user-optimized environment. *EURO Journal on Computational Optimization*, 8:61–84, 2020. doi: https://doi.org/10.1007/s13675-019-00120-w.

Dempe, S. *Foundations of Bilevel Programming*, 61 of *Nonconvex Optimization and Its Applications*. Springer, 2002.

Dempe, S. and Zemkoho, A. *Bilevel Optimization: Advances and Next Challenges*. Springer, 2020.

Deutsch, Y. and Golany, B. A parcel locker network as a solution to the logistics last mile problem. *International Journal of Production Research*, 56(1-2):251–261, 2018. doi: 10.1080/00207543.2017.1395490.

Diakova, Z. and Kochetov, Y. A double VNS heuristic for the facility location and pricing problem. *Electronic Notes in Discrete Mathematics*, 39:29 – 34, 2012. doi: https://doi.org/10.1016/j.endm.2012.10.005.

Drexl, M. and Schneider, M. A survey of variants and extensions of the location-routing problem. *European Journal of Operational Research*, 241(2):283–308, 2015. doi: https://doi.org/10.1016/j.ejor.2014.08.030.

Duong, N. H., Dam, T. T., Ta, T. A., and Mai, T. Joint location and cost planning in maximum capture facility location under random utilities. *Computers & Operations Research*, 159:106336, 2023. doi: 10.1016/j.cor.2023.106336.

Emelogu, A., Chowdhury, S., Marufuzzaman, M., Bian, L., and Eksioglu, B. An enhanced sample average approximation method for stochastic optimization. *International Journal of Production Economics*, 182:230–252, 2016. doi: https://doi.org/10.1016/j.ijpe.2016.08.032.

Espejo, I., Marín, A., and Rodríguez-Chía, A. M. Closest assignment constraints in discrete location problems. *European Journal of Operational Research*, 219(1): 49–58, 2012. doi: 10.1016/j.ejor.2011.12.00.

Görtz, S. and Klose, A. A simple but usually fast branch-and-bound algorithm for the capacitated facility location problem. *INFORMS Journal on Computing*, 4: 597–610, 2011. doi: https://doi.org/10.1287/ijoc.1110.0468.

Grand View Research. Logistics market size, share & growth analysis report, 2030. `www.grandviewresearch.com/industry-analysis/logistics-market-report`, 2023. Accessed: 2025-02-11.

Guerrero-Lorente, J., Gabor, A. F., and Ponce-Cueto, E. Omnichannel logistics network design with integrated customer preference for deliveries and returns. *Computers & Industrial Engineering*, 144:106433, 2020. doi: https://doi.org/10.1016/j.cie.2020.106433.

Gutjahr, W. J. and Dzubur, N. Bi-objective bilevel optimization of distribution center locations considering user equilibria. *Transportation Research Part E: Logistics and Transportation Review*, 85:1 – 22, 2016. doi: https://doi.org/10.1016/j.tre.2015.11.001.

Haase, K. Discrete location planning. technical report wp-09-07. Technical report, Institute for Transport and Logistics Studies, University of Sydney, 2009.

Haase, K. and Müller, S. Management of school locations allowing for free school choice. *Omega*, 41(5):847–855, 2013. doi: 10.1016/j.omega.2012.10.008.

Haase, K. and Müller, S. A comparison of linear reformulations for multinomial logit choice probabilities in facility location models. *European Journal of Operational Research*, 232(3):689–691, 2014. doi: https://doi.org/10.1016/j.ejor.2013.08.009.

Hammami, F., Rekik, M., and Coelho, L. C. An exact method for the combinatorial bids generation problem with uncertainty on clearing prices, bids success, and contracts materialization. *Computers & Operations Research*, 148:105982, 2022. doi: https://doi.org/10.1016/j.cor.2022.105982.

Hansen, P., Jaumard, B., and Savard, G. New branch-and-bound rules for linear bilevel programming. *SIAM Journal on Scientific and Statistical Computing*, 13 (5):1194–1217, 1992. doi: 10.1137/0913069.

Hui Lin, Y., Tian, Q., and Zhao, Y. Unified framework for choice-based facility location problem. *INFORMS Journal on Computing*, 2024. doi: 10.1287/ijoc.2022.0366.

Intelligence, M. Canada freight and logistics market - growth, trends, and forecasts (2023 - 2028), 2024. URL `https://www.mordorintelligence.com/industry-reports/canada-freight-logistics-market-study`. Accessed: 2024-08-28.

Janjevic, M., Winkenbach, M., and Merchán, D. Integrating collection-and-delivery points in the strategic design of urban last-mile e-commerce distribution networks. *Transportation Research Part E: Logistics and Transportation Review*, 131:37–67, 2019. doi: https://doi.org/10.1016/j.tre.2019.09.001.

Kaut, M. and Wallace, S. Evaluation of scenario-generation methods for stochastic programming. *Pacific Journal of Optimization*, 3, 06 2003.

Kleinert, T., Labbé, M., Plein, F., and Schmidt, M. Technical note—there's no free lunch: On the hardness of choosing a correct Big-M in bilevel optimization. *Operations Research*, 68(6):1716–1721, 2020. doi: 10.1287/opre.2019.1944.

Klose, A. CFLP-instance-generator. Mendeley Data, V1, 2018. URL `https://data.mendeley.com/datasets/vhst7gvh7j/1`. Accessed on October 1, 2022.

Klose, A. and Drexl, A. Lower bounds for the capacitated facility location problem based on column generation. *Management Science*, 51(11):1689–1705, 2005. doi: https://doi.org/10.1287/mnsc.1050.0410.

Klose, A. and Görtz, S. A branch-and-price algorithm for the capacitated facility location problem. *European Journal of Operational Research*, 179:1109–1125, 2007. doi: https://doi.org/10.1016/j.ejor.2005.03.078.

Kress, D. and Pesch, E. Competitive location and pricing on networks with random utilities. *Netw Spat Econ*, 16:837–863, 2016. doi: https://doi.org/10.1007/s11067-015-9301-y.

Krohn, R., Müller, S., and Haase, K. Preventive healthcare facility location planning with quality-conscious clients. *OR Spectrum*, 43(1):59–87, 2021. doi: 10.1007/s00291-020-00605-w.

Küçükaydin, H., Aras, N., and Altınel, I. K. Competitive facility location problem with attractiveness adjustment of the follower: A bilevel programming model and its solution. *European Journal of Operational Research*, 208(3):206 – 220, 2011. doi: https://doi.org/10.1016/j.ejor.2010.08.009.

Lafkihi, M., Pan, S., and Ballot, E. Freight transportation service procurement: A literature review and future research opportunities in omnichannel e-commerce. *Transportation Research Part E: Logistics and Transportation Review*, 125:348–365, 2019. doi: https://doi.org/10.1016/j.tre.2019.03.021.

Lamontagne, S., Carvalho, M., Frejinger, E., Gendron, B., Anjos, M. F., and Atallah, R. Optimising electric vehicle charging station placement using advanced discrete choice models. *INFORMS Journal on Computing*, 35(5):1195–1213, 2023. doi: 10.1287/ijoc.2022.0185.

Langley, C. and Infosys. 23rd annual third-party logistics study: The state of logistics outsourcing, 2019. URL `https://www.kornferry.com/content/dam/kornferry/docs/article-migration/2019-3PL-Study.pdf`. Accessed on December 22, 2022.

Lee, J. M. and Lee, Y. H. Facility location and scale decision problem with customer preference. *Computers & Industrial Engineering*, 63(1):184 – 191, 2012. doi: https://doi.org/10.1016/j.cie.2012.02.005.

Legault, R. and Frejinger, E. A model-free approach for solving choice-based competitive facility location problems using simulation and submodularity. *INFORMS Journal on Computing*, 0(0), 2023. doi: 10.1287/ijoc.2023.0280.

Lin, Y. H. and Tian, Q. Facility location and pricing problem: discretized mill price and exact algorithms. *European Journal of Operational Research*, 2022. doi: https://doi.org/10.1016/j.ejor.2022.11.052.

Lin, Y. H., Wang, Y., He, D., and Lee, L. H. Last-mile delivery: Optimal locker location under multinomial logit choice model. *Transportation Research Part E: Logistics and Transportation Review*, 142:102059, 2020. doi: 10.1016/j.tre.2020.102059.

Lin, Y., Wang, Y., Lee, L. H., and Chew, E. P. Profit-maximizing parcel locker location problem under threshold luce model. *Transportation Research Part E: Logistics and Transportation Review*, 157:102541, 2022. doi: https://doi.org/10.1016/j.tre.2021.102541.

Ljubić, I. and Moreno, E. Outer approximation and submodular cuts for maximum capture facility location problems with random utilities. *European Journal of Operational Research*, 266(1):46–56, 2018. doi: https://doi.org/10.1016/j.ejor.2017.09.023.

Logistics, I. 2022 Inbound Logistics Perspectives: 3PL Market Research Report, 2022. URL `https://www.inboundlogistics.com/articles/2022-inbound-logistics-perspectives-3pl-market-research-report/`. Accessed: 2024-08-28.

Lopes, R. B., Ferreira, C., Santos, B. S., and Barreto, S. A taxonomical analysis, current methods and objectives on location-routing problems. *International Transactions in Operational Research*, 20(6):795–822, 2013. doi: https://doi.org/10.1111/itor.12032.

Luo, R., Ji, S., and Ji, Y. An active-learning pareto evolutionary algorithm for parcel locker network design considering accessibility of customers. *Computers & Operations Research*, 141:105677, 2022. doi: https://doi.org/10.1016/j.cor.2021.105677.

Lyu, K., Chen, H., and Che, A. A bid generation problem in truckload transportation service procurement considering multiple periods and uncertainty: Model and benders decomposition approach. *IEEE Transactions on Intelligent Transportation Systems*, 1–14, 2021. doi: https://doi.org/10.1109/TITS.2021.3091692.

Lüer-Villagra, A. and Marianov, V. A competitive hub location and pricing problem. *European Journal of Operational Research*, 231(3):734 – 744, 2013. doi: https://doi.org/10.1016/j.ejor.2013.06.006.

Mai, T. and Lodi, A. A multicut outer-approximation approach for competitive facility location under random utilities. *European Journal of Operational Research*, 284(3):874–881, 2020. doi: https://doi.org/10.1016/j.ejor.2020.01.020.

Mara, S. T. W., Kuo, R., and Asih, A. M. S. Location-routing problem: a classification of recent research. *International Transactions in Operational Research*, 28 (6):2941–2983, 2021. doi: https://doi.org/10.1111/itor.12950.

McFadden, D. L. Econometric models of probabilistic choice. In Manski, C. F. and McFadden, D. L., editors, *Structural Analysis of Discrete Data with Econometric Applications*, 198–272. MIT Press, Cambridge, MA, USA, 1981.

Melvyn Peters, R. C. L. James Cooper and Randall, H. L. The third-party logistics industry in europe: Provider perspectives on the industry's current status and future prospects. *International Journal of Logistics Research and Applications*, 1 (1):9–25, 1998. doi: 10.1080/13675569808962035.

Molin, E., Kosicki, M., and van Duin, R. Consumer preferences for parcel delivery methods: The potential of parcel locker use in the netherlands. *European Journal of Transport and Infrastructure Research*, 22(2):183–200, 2022. doi: 10.18757/ejtir.2022.22.2.6427.

Morganti, E., Seidel, S., Blanquart, C., Dablanc, L., and Lenz, B. The impact of e-commerce on final deliveries: Alternative parcel delivery services in france and germany. *Transportation Research Procedia*, 4:178–190, 2014. doi: https://doi.org/10.1016/j.trpro.2014.11.014.

Méndez-Vogel, G., Marianov, V., and Lüer-Villagra, A. The follower competitive facility location problem under the nested logit choice rule. *European Journal of Operational Research*, 310(2):834–846, 2023. doi: https://doi.org/10.1016/j.ejor.2023.03.008.

Nagy, G. and Salhi, S. Location-routing: Issues, models and methods. *European Journal of Operational Research*, 177(2):649–672, 2007. doi: https://doi.org/10.1016/j.ejor.2006.04.004.

Pacheco Paneque, M., Bierlaire, M., Gendron, B., and Sharif Azadeh, S. Integrating advanced discrete choice models in mixed integer linear optimization. *Transportation Research Part B: Methodological*, 146:26–49, 2021. doi: https://doi.org/10.1016/j.trb.2021.02.003.

Panin, A. A., Pashchenko, M. G., and Plyasunov, A. V. Bilevel competitive facility location and pricing problems. *Automation and Remote Control*, 75(4):715–727, 2014. doi: https://doi.org/10.1134/S0005117914040110.

Pinzon Ulloa, D., Frejinger, E., and Gendron, B. A logistics provider's profit maximization facility location problem with random utility maximizing followers. *Computers & Operations Research*, 167:106649, 2024. doi: https://doi.org/10.1016/j.cor.2024.106649.

Placek, M. Most outsourced services in global transportation and logistics industry 2021. `https://www.statista.com/statistics/660142/transportation-and-logistics-most-outsourced-services`, 2022. Accessed: 2025-02-11.

Prodhon, C. and Prins, C. A survey of recent research on location-routing problems. *European Journal of Operational Research*, 238(1):1–17, 2014. doi: https://doi.org/10.1016/j.ejor.2014.01.005.

Ricard, L. M. and Bierlaire, M. 50 years of behavioral models for transportation and logistics, 2024. URL `https://infoscience.epfl.ch/handle/20.500.14299/205973`.

Rohmer, S. U. and Gendron, B. Cirrelt-2020-11 a guide to parcel lockers in last mile distribution – highlighting challenges and opportunities from an or perspective. 2020.

Sadana, U., Chenreddy, A., Delage, E., Forel, A., Frejinger, E., and Vidal, T. A survey of contextual optimization methods for decision-making under uncertainty. *European Journal of Operational Research*, 320(2):271–289, 2025. doi: https://doi.org/10.1016/j.ejor.2024.03.020.

Saidani, N., Chu, F., and Chen, H. Competitive facility location and design with reactions of competitors already in the market. *European Journal of Operational Research*, 219(1):9–17, 2012. doi: https://doi.org/10.1016/j.ejor.2011.12.017.

Schiffer, M. and Walther, G. An adaptive large neighborhood search for the location-routing problem with intra-route facilities. *Transportation Science*, 52(2):331–352, 2018. doi: 10.1287/trsc.2017.0746.

Shapiro, A. Monte carlo sampling methods. In *Stochastic Programming*, 10 of *Handbooks in Operations Research and Management Science*, 353–425. Elsevier, 2003. doi: https://doi.org/10.1016/S0927-0507(03)10006-0.

Sinha, A., Malo, P., Deb, K., Korhonen, P., and Wallenius, J. Solving bilevel multi-criterion optimization problems with lower level decision uncertainty. *IEEE Transactions on Evolutionary Computation*, 20(2):199–217, 2016. doi: 10.1109/TEVC.2015.2443057.

Stenger, A., Schneider, M., Schwind, M., and Vigo, D. Location routing for small package shippers with subcontracting options. *International Journal of Production Economics*, 140(2):702–712, 2012. doi: https://doi.org/10.1016/j.ijpe.2011.11.014.

Suárez-Vega, R., Santos-Peñate, D., and Dorta-González, P. Discretization and resolution of the (r|X_p)-medianoid problem involving quality criteria. *Top*, 12:111–133, 2004. doi: https://doi.org/10.1007/BF02578927.

Train, K. E. *Discrete Choice Methods with Simulation*. Cambridge University Press, 2nd edition, 2009.

Van Slyke, R. M. and Wets, R. L-shaped linear programs with applications to optimal control and stochastic programming. *SIAM Journal on Applied Mathematics*, 17 (4):638–663, 1969. doi: 10.1137/0117061.

Vanderbei, R. J. *Linear Programming: Foundations and Extensions*, 196 of *International Series in Operations Research & Management Science*. Springer, 5th edition, 2020. doi: 10.1007/978-3-030-38766-7.

Wolsey, L. A. *Integer Programming*. John Wiley & Sons, 1998.

Xu, J., Tu, Y., and Zeng, Z. A nonlinear multiobjective bilevel model for minimum cost network flow problem in a large-scale construction project. *Mathematical Problems in Engineering*, 2012, 04 2012. doi: 10.1155/2012/463976.

Yan, F., Ma, Y., and Feng, C. A bi-level programming for transportation services procurement based on combinatorial auction with fuzzy random parameters. *Asia Pacific Journal of Marketing and Logistics*, 30:1162–1182, 2018. doi: https://doi.

org/10.1108/APJML-07-2017-0154.

Yuen, K. F., Wang, X., Ng, L. T. W., and Wong, Y. D. An investigation of customers' intention to use self-collection services for last-mile delivery. *Transport Policy*, 66: 1–8, 2018. doi: https://doi.org/10.1016/j.tranpol.2018.03.001.

Zhang, Y., Berman, O., and Verter, V. The impact of client choice on preventive healthcare facility network design. *OR Spectrum*, 34(2):349–370, 2012. doi: 10.1007/s00291-011-0280-1.

Zhou, L., Wang, X., Ni, L., and Lin, Y. Location-routing problem with simultaneous home delivery and customer's pickup for city distribution of online shopping purchases. *Sustainability*, 8(8), 2016. doi: 10.3390/su8080828.

Appendix A

First Article

A.1. Notation

Indices	
i	Facility location
m	Service level
n	Shipper
j	Customer
k	Customer category
p	Price level
n_j	Shipper associated to customer j
k_j	Customer category associated to customer j
Sets	
I	Facility locations
M	Service levels defined by the logistics provider
N	Shippers
J	Customers
K	Categories of customers
P	Price levels offered by the logistics provider
P_n^m	Price levels available to shipper $n \in N$ for a service level $m \in M$
J_k	Customers in customer category $k \in K$, $(J_k \subset J)$
$J_n^{'}$	Customers that belong to shipper $n \in N$, $(J_n^{'} \subset J)$
K_n	Customer categories of shipper $n \in N$
M_n	Services levels available for shipper $n \in N$
M_{nk}	Services levels available for shipper $n \in N$ and customers in category $k \in K_n$
M_{nk}^0	$M_{nk} \cup \{0\}$ (includes the opt-out alternative)
M_j	$\{m \in M : \exists k \in K$ such that $j \in J_k$ and $m \in M_k\}$
$F_{nk}(\mathbf{y},\mathbf{z})$	Set of optimal solutions for the follower's problem for each n and k given y_n^{mp} and z_{nk}^m
Parameters	
c_{ij}^m	Cost of assigning customer j to facility location i according to service level m
u_i	Capacity of facility i
f_i	Fixed cost for facility location $i \in I$
d_j	Demand corresponding to customer j
d_k	$\sum_{j \in J_k} d_j$
γ^m	Scale factor of capacity usage given by service level m
l_n^{mp}	Minimum demand required to shipper n for price level p and service level m
q_n^{mp}	Price level p for service level m available for the shipper n
L_{nk}^m	Preference of customers of shipper n in category k to service level m

Table A.1. Indexes, sets and parameters

A.2. Single-level MILP

Here, we present the single-level formulation that we solve with a standard Benders decomposition algorithm. As shown in Proposition 1, the follower's problem is linear. Hence, after replacing constraints (2.8f) with their corresponding single-level reformulation obtained from applications of strong duality theorem and standard linearization techniques, the resulting single-level MILP formulation is as follows:

Stochastic data representation

$\boldsymbol{\varepsilon}$	Random vector
ε_{nk}^{m}	Random variable in the utility of shipper n and customers in k for service level m
S	Set of scenarios of ε
S_{nk}^{m}	Set of scenarios where $U_{nks}^{m}(\mathbf{y};\varepsilon_{nks}^{m}) > U_{nk}^{0}(\varepsilon_{nks}^{0})$ for each n, k and m
s	Index for scenario
$\boldsymbol{\varepsilon_s}$	Random realization of $\boldsymbol{\varepsilon}$ for scenario s
ε_{nks}^{m}	Random realization of ε_{nk}^{m} for scenario s
$F_{nk}(\mathbf{y},\mathbf{z};\boldsymbol{\varepsilon})$	$F_{nk}(\mathbf{y},\mathbf{z})$ given $\boldsymbol{\varepsilon}$
$F_{nks}(\mathbf{y},\mathbf{z};\boldsymbol{\varepsilon_s})$	$F_{nk}(\mathbf{y},\mathbf{z})$ for scenario, $s \in S$
ρ_{nk}^{mp}	Probability that $U_{nk}^{m}(\mathbf{y};\varepsilon_{nk}^{m}) > U_{nk}^{0}(\varepsilon_{nk}^{0})$

Decision and auxiliary variables

w_{ij}^{m}	Fraction of the demand d_j assigned to facility i with a service level m
r_i	1 If facility i is implemented; 0 otherwise.
t_n^m	Price assigned to service level m chosen for shipper n
y_n^{mp}	$\begin{cases} 1 & \text{if LP offers price level } p \text{ and service level } m \text{ to shipper } n \\ 0 & \text{otherwise} \end{cases}$
z_{nk}^{m}	$\begin{cases} 1 & \text{if LP assigns service level } m \text{ to shipper } n \text{ and customer category } k \\ 0 & \text{otherwise} \end{cases}$
x_{nk}^{m}	$\begin{cases} 1; & \text{if shipper } n \text{ accepts service level } m \text{ for the customer category } k \\ 0; & \text{otherwise} \end{cases}$
π_{nk}^{mp}	Auxiliary variable to linearize $z_{nk}^{m} y_n^{mp}$
ν_{ij}^{mp}	Auxiliary variable to linearize $w_{ij}^{m} y_n^{mp}$
$\mathbf{y}$	vector representation of decision variables y_n^{mp}
$\mathbf{z}$	vector representation of decision variables z_{nk}^{m}
$\mathbf{x}$	vector representation of decision variables x_{nk}^{m}
$\mathbf{r}$	vector representation of decision variables r_i
$\mathbf{w}$	vector representation of decision variables w_{ij}^{m}

Functions

$U_n^m(\mathbf{y})$	Deterministic utility of service level m for shipper n, given y_n^{mp}
$U_{nk}^{m}(\mathbf{y})$	Deterministic utility of service level m, shipper n and category k, given y_n^{mp}
U_{nk}^{0}	Deterministic utility of opt-out alternative for shipper n and category k
$U_{nks}^{m}(\mathbf{y};\varepsilon_{nks}^{m})$	$U_{nk}^{m}(y)$ for scenario $s \in S$
$U_{nks}^{0}(\varepsilon_{nks}^{0})$	$U_{nk}^{0}(y)$ for scenario $s \in S$
$Q(\mathbf{r},\mathbf{z},\mathbf{y})$	Recourse function
$\bar{Q}(\mathbf{r},\mathbf{z},\mathbf{y})$	SAA $Q(\mathbf{r},\mathbf{z},\mathbf{y})$
$\phi(\mathbf{r},\mathbf{z},\mathbf{y},\boldsymbol{\varepsilon})$	Second stage problem given random variables $\boldsymbol{\varepsilon}$
$\phi_s(\mathbf{r},\mathbf{z},\mathbf{y},\boldsymbol{\varepsilon_s})$	Second stage problem for scenario $s \in S$

Table A.2. Stochastic data, variables and functions for model formulation

$$\max -\sum_{i\in I} f_i r_i + \frac{1}{|S|}\Bigg(\sum_{n\in N}\sum_{k\in K_n}\sum_{m\in M_k}\sum_{p\in P_n^m} d_k q_n^{mp}\pi_{nks}^{mp} - \sum_{i\in I}\sum_{j\in J}\sum_{m\in M_j} c_{ij}^m d_j w_{ijs}^m\Bigg) \tag{A.1}$$

s.t. (2.6b) - (2.6h), (2.8b) - (2.8e), (2.8g)

$$\sum_{m\in M_{nk}^0}\left(\alpha\sum_{p\in P_n^m} q_n^{mp}\pi_{nks}^{mp}\right) + L_{nks}^m x_{nks}^m \geq$$

$$\lambda_{nks} + \sum_{m\in M_{nk}} \theta_{nks}^m, \qquad n\in N, k\in K_n, s\in S, \tag{A.2}$$

$$\sum_{m\in M_{nk}^0} x_{nks}^m = 1, \qquad n\in N, k\in K_n, s\in S, \tag{A.3}$$

$$x_{nks}^m \leq z_{nk}^m, \qquad n\in N, k\in K_n, m\in M_{nk}, s\in S, \tag{A.4}$$

$$\lambda_{nks} + \delta_{nks}^m - \alpha\sum_{p\in P_n^m} q_n^{mp} y_n^{mp} \geq L_{nks}^m, \qquad n\in N, k\in K_n, m\in M_{nk}, s\in S, \tag{A.5}$$

$$\lambda_{nks} \geq U_{nks}^0, \qquad n\in N, k\in K_n, s\in S, \tag{A.6}$$

$$\pi_{nks}^{mp} \leq x_{nks}^m, \qquad n\in N, k\in K_n, m\in M_{nk}, p\in P_n^m, s\in S, \tag{A.7}$$

$$\pi_{nks}^{mp} \leq y_n^{mp}, \qquad n\in N, k\in K_n, m\in M_{nk}, p\in P_n^m, s\in S, \tag{A.8}$$

$$\pi_{nks}^{mp} \geq x_{nks}^m - 1 + y_n^{mp}, \qquad n\in N, k\in K_n, m\in M_{nk}, p\in P_n^m, s\in S, \tag{A.9}$$

$$\pi_{nks}^{mp} \geq 0, \qquad n\in N, k\in K_n, m\in M_{nk}, p\in P_n^m, s\in S, \tag{A.10}$$

$$\theta_{nks}^m \leq \delta_{nks}^m, \qquad n\in N, k\in K_n, m\in M_{nk}, s\in S, \tag{A.11}$$

$$\theta_{nks}^m \leq \overline{U}_{nks} z_{nk}^m, \qquad n\in N, k\in K_n, m\in M_{nk}, s\in S, \tag{A.12}$$

$$\theta_{nks}^m \geq \delta_{nks}^m - \overline{U}_{nks}(1 - z_{nk}^m), \qquad n\in N, k\in K_n, m\in M_{nk}, s\in S, \tag{A.13}$$

$$\theta_{nks}^m \geq 0, \qquad n\in N, k\in K_n, m\in M_{nk}, s\in S, \tag{A.14}$$

$$\delta_{nks}^m \geq 0, \qquad n\in N, k\in K_n, m\in M_{nk}, s\in S. \tag{A.15}$$

where $\overline{U}_{nks}$ denotes a large value, in this case, the maximum value obtained for U_{nks}^{mp}, as stated in Proposition 6 below. The sets of constraints (A.2), (A.3), (A.4), (A.5) and (A.6) represent the single-level reformulation resulting from the strong duality theorem. There, (A.3) and (A.4) are the sets of constraints in the follower's primal problem, while (A.5) and (A.6) the sets of constraints in the corresponding dual. Moreover, λ_{nks} and δ_{nks}^m are the corresponding dual variables, associated with the primal constraints (A.3) and (A.4), respectively. Constraints (A.7) - (A.10)

and constraints (A.11) - (A.14) are required to linearize $\pi^{mp}_{nks} = x^m_{nks} y^{mp}_n$ and $\theta^m_{nks} = \delta^m_{nks} z^m_{nk}$.

Proposition 6. *$\overline{U}_{nks} = \max\{U^{mp}_{nks}, m \in M_{nk}, p \in P^m_n\}$ is a valid bound for δ^m_{nks}.*

Proof. Note that $U^m_{nks} = \alpha \sum_{p \in P^m_n} q^{mp}_n y^{mp}_n + L^m_{nks}$. Given that $\sum_{p \in P^m_n} y^{mp}_n = 1$, $U^m_{nks} \leq \overline{U}_{nks}, \forall m \in M_{nk}$. Therefore, the dual constraints (A.5) and (A.6) for each n, k, s can be respectively written as $\lambda_{nks} + \delta^m_{nks} \geq U^m_{nks}, \forall m \in M_{nk}$ and $\lambda_{nks} \geq U^0_{nks}$. Hence, as in the dual we minimize the objective function given by $\lambda_{nks} + \sum_{m \in M_{nk}} \delta^m_{nks} z^m_{nk}$, then $\lambda_{nks} + \delta^m_{nks}$ cannot exceed $\max\{U^{mp}_{nks}, \forall m \in M_{nk}, p \in P^m_n\}$. To show this, let us distinguish two cases. First, when $U^0_{nks} > U^m_{nks}, \forall m \in M_{nk}$ the minimum is obtained by $\lambda_{nks} = U^0_{nks}$, in this case $\delta^m_{nks} = 0 < \overline{U}_{nks}$. Second, for $\exists m \in M_{nk}, p \in P^m_n$, such that $U^{mp}_{nks} > U^0_{nks}$, the optimal solution is $\lambda_{nks} = U^0_{nks}$ and $\delta^m_{nks} = U^{mp}_{nks} - U^0_{nks} < \overline{U}_{nks} - U^0_{nks} < \overline{U}_{nks}$. □

A.3. Parameters in Experimental Results

Here, we describe the procedure we use to determine some of the parameter values in the computational experiments. As seen in Section 2.7.3.1, values must be assigned to parameters α and β. For the case of parameter α, we seek for a value that leads to a very low probability ρ^{mp}_{nk} for the first service level ($m = 1$) and lowest price level ($p = 0$,$q^0 = 15$). We set $\rho^{10}_{nk} = 0.005$ to compute the initial value of α. Therefore, we obtain the first value $\alpha = -0.45289$ by solving:

$$\frac{1}{1 + e^{(U^0 - U^{10})}} = \frac{1}{1 + e^{(3 - 15\alpha - 4.5)}} = 0.005.$$

The other 10 values taken by α are equally spaced between the first one and 0. The 11 resulting values for α are reported in Table A.3.

To determine values for parameter β, we use the exponential expression 2^l over the set $l = \{-5, -4, -3, -2, -1, 0, 1, 2, 3\}$. Low values of l correspond to low values for the scale parameter β. Table A.4 reports values for β that are used in Section 2.7.3.1.

Instance
$\alpha = -0.45289$
$\alpha = -0.4076$
$\alpha = -0.36231$
$\alpha = -0.31702$
$\alpha = -0.27173$
$\alpha = -0.22644$
$\alpha = -0.18115$
$\alpha = -0.13587$
$\alpha = -0.09058$
$\alpha = -0.04529$
$\alpha = 0.0$

Table A.3. Possible values of the configuration parameter α

l	β
-5	0.03125
-4	0.0625
-3	0.125
-2	0.25
-1	0.5
0	1
1	2
2	4
3	8

Table A.4. Values for the scale parameter β

Regarding the instances considered in the *ORlib* class, the first group of instances is created by selecting $|P| \in \{5,10,15\}$, $|I| \in \{16,25,50\}$, $|J| = 50$, $u_i \in \{5000,15000, 58268\}$ and $f_i \in \{7500, 12500, 17500, 25000\}$. Hence, $|P| \times |I| \times |u_i| \times |f_i|$ leads to 96 problems. Another group of 4 instances is created with $|I| = 100$, $|J| = 1000$, $u_i \in \{8000, 10000, 12000, 14000\}$, and $|P| = 5$. Except for the price level sets, the rest of the values are the same as in Beasley (1988).Table A.5 shows the sets of price levels for the *Orlib* class when $|P| = 5, |P| = 10, |P| = 15$ for each service level.

Service Level	$\|P\|$	Price Level Set
0	5	{11000,17000,23000,29000,35000}
1	5	{12000,18000,24000,30000,36000}
2	5	{13000,19000,25000,31000,37000}
0	10	{11000,14000,17000,20000,23000,26000,29000,32000,35000,38000}
1	10	{12000,15000,18000,21000,24000,27000,30000,33000,36000,39000}
2	10	{13000,16000,19000,22000,25000,28000,31000,34000,37000,40000}
0	15	{2000,5000,8000,11000,14000,17000,20000,23000,26000,29000,32000,35000,38000,41000,44000}
1	15	{3000,6000,9000,12000,15000,18000,21000,24000,27000,30000,33000,36000,39000,42000,45000}
2	15	{4000,7000,10000,13000,16000,19000,22000,25000,28000,31000,34000,37000,40000,43000,46000}

Table A.5. Price level sets for the *ORlib* class of instances

A.4. Proposition in Section 2.5.2

Proposition 7. *The cost in objective function* (2.8a)*,*

$$\sum_{m\in M_j}\sum_{i\in I}\sum_{j\in J}\left(c_{ij}^m d_j w_{ijs}^m\right),$$

can be written as

$$\sum_{n\in N}\sum_{k\in K_n}\sum_{m\in M_{nk}}\sum_{i\in I}\sum_{j\in J_n'\cap J_k}\left(c_{ij}^m d_j w_{ijs}^m \sum_{p\in P_n^m} y_n^{mp}\right).$$

Proof. A customer $j \in J$ belongs to a unique shipper n and customer category k, that is $J = \bigcup_{n\in N,k\in K_n} J_n^{'} \cap J_k$.Also, the set of service levels M_j available for a customer j corresponds to the sets of service levels available for the shipper n and customer category k where it belongs, that is $M_j = \{m \in M_{nk} : j \in J_n^{'} \cap J_k\}$. Therefore, it follows that:

$$\sum_{m\in M_j}\sum_{i\in I}\sum_{j\in J}\left(c_{ij}^m d_j w_{ijs}^m\right) = \sum_{n\in N}\sum_{k\in K_n}\sum_{m\in M_{nk}}\sum_{i\in I}\sum_{j\in J_n'\cap J_k}\left(c_{ij}^m d_j w_{ijs}^m\right). \tag{A.16}$$

Let us assume given n and m. Then, for each $j \in J_n' \cap J_k$, multiply w_{ijs}^m by $\sum_{p\in P_n^m} y_n^{mp}$. Recall that and $\sum_{p\in P_n^m} y_n^{mp} \leq 1$ and $y_n^{mp} \in \{0,1\}$, as stated in constraints (2.6b), (2.6g). If $\sum_{p\in P_{n^\star}^m} y_n^{mp} = 0$, then $z_{nk}^m = 0$, $k \in K_n$ due to constraints (2.6e), $x_{nks}^m = 0$ because $x_{nks}^m \leq z_{nk}^m$, as seen in (2.9), and $w_{ijs}^m = 0$ for $i \in I, j \in J_k$, due to constraints (2.8d). Hence, $\sum_{i\in I}\sum_{j\in J_n'\cap J_k} c_{ij}^m d_j w_{ijs}^m = 0$. Also

$\sum_{k\in K_n}\sum_{i\in I}\sum_{j\in J'_n\cap J_k}\left(c^m_{ij}d_j w^m_{ijs}\sum_{p\in P^m_n} y^{mp}_n\right)=0$. On the contrary, if $\sum_{p\in P^m_n} y^{mp}_n = 1$, then $\sum_{i\in I}\sum_{j\in J'_n\cap J_k} c^m_{ij}d_j w^m_{ijs}\sum_{p\in P^m_n} y^{mp}_n = \sum_{i\in I}\sum_{j\in J'_n\cap J_k} c^m_{ij}d_j w^m_{ijs}$. □

A.5. Detailed Results

The relative gap for the instances with zero profit is 0%. In Table A.6 we note that 41 cases correspond to instance configurations where the optimal profit obtained by *BA* is far from stability (a relative difference larger than 20%). Table A.7 shows that *H-S* and *H-A* clearly outperform *BA*. Table A.8 shows that *H-S* and *H-A* provide high quality solutions. Recall, however, that *H-A* utilizes analytical probabilities instead of estimates computed from simulated utilities as in *BA* and *H-S*.

			$\frac{Z_{90\%}-Z_{10\%}}{Z_{10\%}}$			
$\|I\|-\|J\|$	U^0	β	$\|S\|=100$	$\|S\|=500$	$\|S\|=1000$	$\|S\|=5000$
2-15	0	0.03125	0%	0%	0%	0%
2-15	0	0.0625	0%	0%	0%	0%
2-15	0	0.125	0%	0%	0%	0%
2-15	0	0.25	1.4%	0.5%	0.3%	0.2%
2-15	0	0.5	4.6%	1.6%	2.2%	1.2%
2-15	0	1	11.3%	5.8%	4.0%	2.3%
2-15	0	2	24.4%	24.7%	9.3%	4.8%
2-15	0	4	80%	48.4%	22.2%	8.7%
2-15	0	8	159.5%	83.5%	47.6%	**20.8%**
2-15	0	16	432.9%	205.7%	80.1%	**39.4%**
2-15	0	32	1022.2%	4655.4%	271.0%	**53.3%**
2-15	4.5	0.03125	0%	0%	0%	0%
2-15	4.5	0.0625	0%	0%	0%	0%
2-15	4.5	0.125	0%	0%	0%	0%
2-15	4.5	0.25	0%	0%	0%	0%
2-15	4.5	0.5	0%	0%	0%	0%
2-15	4.5	1	0%	0%	0%	0%
2-15	4.5	2	0%	0%	0%	0%
2-15	4.5	4	0%	0%	0%	0%
2-15	4.5	8	0%	0%	0%	0%
2-15	4.5	16	-	0%	0%	0%
2-15	4.5	32	-	-	0%	0%
2-30	0	0.03125	0%	0%	0%	0%
2-30	0	0.0625	0%	0%	0%	0%
2-30	0	0.125	0%	0%	0%	0%
2-30	0	0.25	1.2%	0.4%	0.3%	0.1%
2-30	0	0.5	4.5%	1.4%	1.8%	1.0%
2-30	0	1	8.5%	4.9%	3.6%	1.8%
2-30	0	2	14.8%	15.4%	5.3%	3.1%
2-30	0	4	31.6%	21.1%	9.0%	4.1%
2-30	0	8	41.6%	20.6%	12.6%	5.6%
2-30	0	16	54.0%	25.8%	12.5%	7.0%
2-30	0	32	56.5%	29.8%	15.9%	6.5%
2-30	4.5	0.03125	0%	0%	0%	0%
2-30	4.5	0.0625	0%	0%	0%	0%
2-30	4.5	0.125	0%	0%	0%	0%
2-30	4.5	0.25	0%	0%	0%	0%
2-30	4.5	0.5	0.0%	0%	0%	0%
2-30	4.5	1	0.0%	0%	0%	0%
2-30	4.5	2	0.0%	0%	0%	0%
2-30	4.5	4	-	0%	0%	0%
2-30	4.5	8	409.5%	510.8%	167.0%	**62.7%**
2-30	4.5	16	107.6%	71.4%	35.1%	11.5%
2-30	4.5	32	79.7%	39.0%	17.5%	9.4%
2-45	0	0.03125	0%	0%	0%	0%
2-45	0	0.0625	0.0%	0.0%	0.0%	0%
2-45	0	0.125	0.0%	0.0%	0.0%	0.0%
2-45	0	0.25	1%	0.4%	0.3%	0.1%
2-45	0	0.5	3.7%	1.3%	1.5%	0.9%
2-45	0	1	7.1%	3.7%	2.7%	1.5%
2-45	0	2	12.4%	11.0%	4.4%	2.2%
2-45	0	4	22.9%	14.3%	6.9%	3.0%
2-45	0	8	28.2%	14.2%	8.3%	3.6%
2-45	0	16	34.1%	15.7%	7.8%	4.3%
2-45	0	32	34.5%	17.9%	9.6%	3.6%
2-45	4.5	0.03125	0%	0%	0%	0%
2-45	4.5	0.0625	0%	0%	0%	0%
2-45	4.5	0.125	0%	0%	0%	0%
2-45	4.5	0.25	0%	0%	0%	0%
2-45	4.5	0.5	0%	0%	0%	0%
2-45	4.5	1	0%	0%	0%	0%
2-45	4.5	2	0%	0%	0%	0%
2-45	4.5	4	-	1126.9%	180.7%	**84.4%**
2-45	4.5	8	48.6%	37.5%	20%	8.4%
2-45	4.5	16	46.1%	27.9%	14.8%	4.6%
2-45	4.5	32	42.7%	19.8%	8.8%	4.2%

Table A.6. Relative differences between $Z_{10\%}$ and $Z_{90\%}$ for optimal solutions with *BA*

			Avg. Computing Time (sec.)		Computing Time (sec.)
$\|I\| - \|J\|$	U^0	β	***BA***	***H-S***	***H-A***
2-15	0	0.03125	1244.4	0.0	0.1
2-15	0	0.0625	1974.0	0.1	0.1
2-15	0	0.125	3082.0	0.0	0.1
2-15	0	0.25	4083.0	0.1	0.1
2-15	0	0.5	4317.0	0.0	0.1
2-15	0	1	8197.0	0.0	0.1
2-15	0	2	7735.0	0.1	0.1
2-15	0	4	4195.0	0.1	0.1
2-15	4.5	0.03125	0.1	0.0	0.0
2-15	4.5	0.0625	0.1	0.0	0.0
2-15	4.5	0.125	3.0	0.0	0.0
2-15	4.5	0.25	4.7	0.0	0.0
2-15	4.5	0.5	5.8	0.0	0.0
2-15	4.5	1	11.7	0.0	0.0
2-15	4.5	2	157.7	0.0	0.0
2-15	4.5	4	672.2	0.0	0.0
2-15	4.5	8	1167.3	0.0	0.0
2-15	4.5	16	1642.0	0.0	0.1
2-15	4.5	32	1479.0	0.1	0.1
2-30	0	0.03125	1317.8	0.1	0.2
2-30	0	0.0625	2007.0	0.1	0.2
2-30	0	0.125	2607.0	0.1	0.2
2-30	0	0.25	3870.0	0.1	0.2
2-30	0	0.5	4252.0	0.1	0.2
2-30	0	1	7085	0.1	0.2
2-30	0	2	6118.0	0.1	0.2
2-30	0	4	4222.0	0.1	0.2
2-30	0	8	2689.0	0.1	0.2
2-30	0	16	1897.0	0.1	0.2
2-30	0	32	1703.0	0.1	0.2
2-30	4.5	0.03125	0.2	0.0	0.0
2-30	4.5	0.0625	0.2	0.0	0.0
2-30	4.5	0.125	1.5	0.0	0.0
2-30	4.5	0.25	4.0	0.0	0.0
2-30	4.5	0.5	4.7	0.0	0.0
2-30	4.5	1	21.8	0.0	0.0
2-30	4.5	2	245.7	0.0	0.1
2-30	4.5	4	1388.3	0.1	0.2
2-30	4.5	16	1738.0	0.1	0.2
2-30	4.5	32	1582.0	0.1	0.2
2-45	0	0.03125	876.5	0.1	0.2
2-45	0	0.0625	1671.7	0.1	0.2
2-45	0	0.125	2665.0	0.1	0.2
2-45	0	0.25	3546.0	0.2	0.3
2-45	0	0.5	3863.0	0.2	0.3
2-45	0	1	5048.0	0.2	0.3
2-45	0	2	5898.0	0.2	0.3
2-45	0	4	3098.0	0.2	0.3
2-45	0	8	2296.0	0.2	0.3
2-45	0	16	1653.0	0.2	0.3
2-45	0	32	1418.0	0.2	0.3
2-45	4.5	0.03125	0.2	0.0	0.0
2-45	4.5	0.0625	0.2	0.0	0.0
2-45	4.5	0.125	1.6	0.0	0.0
2-45	4.5	0.5	4.9	0.0	0.0
2-45	4.5	1	33.6	0.0	0.0
2-45	4.5	2	442.8	0.0	0.0
2-45	4.5	4	1289.0	0.2	0.2
2-45	4.5	8	1594.0	0.2	0.2
2-45	4.5	16	1325.0	0.2	0.3
2-45	4.5	32	1194.0	0.2	0.3

Table A.7. Average computing times for *BA* and *H-S* and *H-A* for each instance configuration

$\|I\| - \|J\|$	U^0	β	**Avg. Gap % *BA* and *H-S***	**Avg. Gap % *BA* and *H-A***
2-15	0	0.03125	0.0	0.0
2-15	0	0.0625	0.0	0.0
2-15	0	0.125	0.0	0.0
2-15	0	0.25	0.0	0.0
2-15	0	0.5	0.1	0.1
2-15	0	1	0.6	1.2
2-15	0	2	1.5	2.6
2-15	0	4	3.0	6.1
2-30	0	0.03125	0.0	0.0
2-30	0	0.0625	0.0	0.0
2-30	0	0.125	0.0	0.0
2-30	0	0.25	0.0	0.0
2-30	0	0.5	0.1	0.1
2-30	0	1	0.3	0.6
2-30	0	2	0.6	1.2
2-30	0	4	0.8	2.1
2-30	0	8	1.1	3.3
2-30	0	16	1.2	3.9
2-30	0	32	1.4	3.9
2-30	4.5	16	2.5	7.1
2-30	4.5	32	1.9	5.2
2-45	0	0.03125	0.0	0.0
2-45	0	0.0625	0.0	0.0
2-45	0	0.125	0.0	0.0
2-45	0	0.25	0.0	0.0
2-45	0	0.5	0.0	0.0
2-45	0	1	0.1	0.4
2-45	0	2	0.1	0.6
2-45	0	4	0.2	1.2
2-45	0	8	0.2	1.8
2-45	0	16	0.2	2.0
2-45	0	32	0.2	2.0
2-45	4,5	8	0.4	4.0
2-45	4,5	16	0.3	2.7
2-45	4,5	32	0.3	2.3

Table A.8. Average optimality gaps between *BA* and *H-S* and *H-A*

Appendix B

Second Article

B.1. Notation

Indexes	
d	CDP
z	Zone
a	Subzone
k	Customer category
q	pattern index
l	capacity level
Sets	
A	Set of Subzones
A_z	Set of subzones in zone z
Z	Set of Zones
K	Set of Customer categories
L	Set of Capacity levels
D	Set of CDPs
D^{P}	Set of pick up points or stores
D^{M}	Set of modular parcel lockers
D^{R}	Set of regular parcel lockers
Q	Set of Patterns
Deterministic parameters	
c_z^{R}	Estimated routing cost for zone z
$c_{dz}^{\tilde{\mathrm{A}}\mathrm{R}}$	Estimated routing cost for zone z affected by CDP d
u_d	Capacity of CDP d
f_d	Fixed cost of CDP d
u_d^l	Capacity level l of CDP d
f_d^l	Fixed cost associated to capacity level l of CDP d
$\underline{b}_d$	Minimum demand level required for CDP $d \in D^{\mathrm{P}}$
$\bar{b}_z$	Average number of parcels to deliver in zone z
b_{az}	Average number of parcels to deliver in subzone a
θ_a	Coordinates of the centroid of subzone a
θ_d	Coordinates of the location of CDP d
b_{zak}	average number of parcels of customers in category k from subzone a in zone z
α_k^1	Distance-related coefficient in the utility function
α_k^2	Coefficient in the utility function related to exogenous characteristics
e_{gdask}	1, If $U_{gaks} \geq U_{daks}$; 0, otherwise

Table B.1. Indexes, sets and deterministic parameters

Derived variables	
$\psi_{dak}(\boldsymbol{x})$	Probability that CDP d captures demand from customers in category k and subzone a
$\psi_{dz}(\boldsymbol{x})$	Probability that CDP d captures demand from zone z
$\hat{\psi}_{dak}(\boldsymbol{x})$	SAA of $\psi_{dak}(\boldsymbol{x})$
ν_{qak}	Fraction of customers in subzone a and category k with the same preference pattern q
U_{dak}	Random utility for CDP d an customers in category k and subzone a
U_{daks}	Realization of U_{dak} for scenario s
ε_{dak}	Random term in the utility function
ε_s	Realization of ε_{dak} in scenario s
Decision variables	
x_d	1, If CDP d is open; 0, otherwise
w_{daks}	1, If customers in category k and subzone a are assigned to CDP d in scenario s; 0, otherwise
p_{dz}	Effective fraction of demand in zone z captured by CDP d given the capacity restrictions
p_{0z}	Effective fraction of demand in zone z captured by or assigned to home delivery service
w_{qd}	1; If demand in pattern q is assigned to CDP d; 0, otherwise
r_d^l	1, If capacity level ld is assigned to CDP d; 0, otherwise

Table B.2. Derived and decision variables

B.2. Dominated Single-level Reformulation

This section introduces the MILP model resulting from reformulating (4.6) by using linear programming strong duality, a similar reasoning as in Proposition 5, and standard linearization techniques:

$$\min \sum_{z\in Z} c_z^{\mathrm{R}} b_z p_{0z} + \sum_{z\in Z}\sum_{d\in D} c_{dz}^{\mathrm{AR}} b_z p_{dz} + \sum_{d\in D^{\mathrm{M}}}\sum_{l\in L_d} f_d^l r_d^l + \sum_{d\in D^{\mathrm{A}}} f_d x_d \tag{B.1a}$$

s.t: $(4.4b - 4.4e, 4.4h - 4.4k)$

$$p_{dz} \leq \frac{1}{|S|}\sum_{s\in S}\sum_{a\in A_z}\sum_{k\in K}\frac{b_{zak}}{b_z} v_{daks}, \qquad d\in D, z\in Z, \tag{B.1b}$$

$$p_{0z} \geq \frac{1}{|S|}\sum_{s\in S}\sum_{a\in A_z}\sum_{k\in K}\frac{b_{zak}}{b_z} v_{0aks}, \qquad z\in Z, \tag{B.1c}$$

$$\sum_{g\in D} U_{gaks} v_{gaks} + U_{0aks} w_{0aks} \geq U_{daks} x_d, \quad d\in D, a\in A, k\in K, s\in S, \tag{B.1d}$$

$$\sum_{g\in D} U_{gaks} v_{gaks} + U_{0aks} w_{0aks} \geq U_{0aks}, \qquad a\in A, k\in K, s\in S, \tag{B.1e}$$

$$\sum_{d\in D\cup\{0\}} w_{daks} = 1, \qquad a\in A, k\in K, s\in S, \tag{B.1f}$$

$$v_{daks} \leq x_d, \qquad d\in D, a\in A, k\in K, s\in S, \tag{B.1g}$$

$$v_{daks} \leq w_{daks}, \qquad d\in D, a\in A, k\in K, s\in S \tag{B.1h}$$

$$v_{daks} \geq x_d + w_{daks} - 1, \qquad d\in D, a\in A, k\in K, s\in S, \tag{B.1i}$$

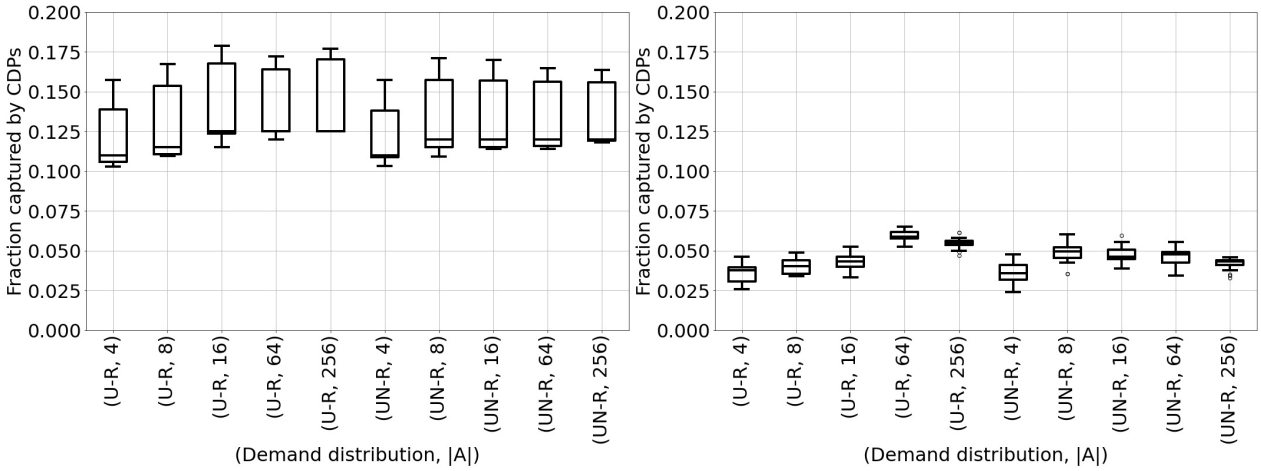
$$v_{daks} \geq 0, \qquad d\in D, a\in A, k\in K, s\in S. \tag{B.1j}$$

Note that this formulation differs from model formulation (4.8) in Constraints (B.1d), (B.1e), (B.1g) - (B.1i). Constraints (B.1d) and (B.1e) force the assignment to the delivery service with the highest utility. Constraints (B.1g) - (B.1i) allow to linearize $v_{daks} = w_{adks} x_d$.

We use two approaches to solve model (B.1), the fist, denoted by D-BB, applies branch-and-bound whereas the second, denoted by D-B, applies Benders decomposition.

B.3. Additional Results

This section provides additional results about performance and optimal solutions. Table B.3 reports the computing times for D-B, D-BB and C-AP, for a subset of instances used in 4.7.2. The panels of Figure B.1 show the distribution of the effective fraction captured by CDPs. Finally, the panels of Figure B.2 show the distribution of the relative performance of C-AP and C-UP with respect to AC-AP.

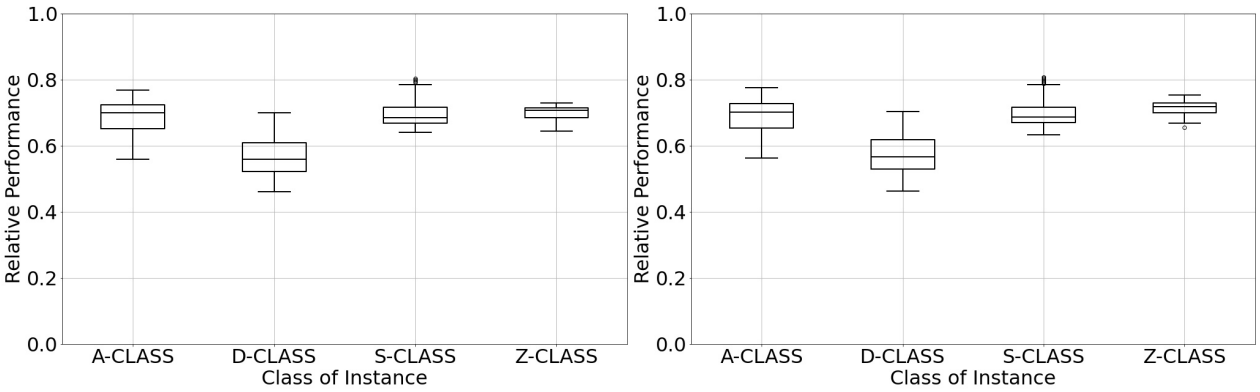


(a) Medium entropy **(b)** Low entropy

Fig. B.1. Distribution of optimal fraction captured by CDPs

Distribution	$\|A\|$	u_d	**Avg. Entropy**	**D-BB**	**D-B**	**C-AP**
U-R	4	20	5.6	1.1	1.1	0.1
U-R	4	20	2.9	2.5	5.9	0.0
U-R	4	20	0.7	0.5	0.8	0.0
U-R	4	30	5.6	1.7	1.9	0.1
U-R	4	30	2.9	3.6	27.8	0.1
U-R	4	30	0.7	0.4	1.0	0.0
U-R	8	20	6.2	3.2	3.0	0.2
U-R	8	20	3.4	7.3	14.5	0.1
U-R	8	20	0.9	2.8	5.7	0.0
U-R	8	30	6.2	4.3	4.9	0.3
U-R	8	30	3.4	8.4	31.3	0.1
U-R	8	30	0.9	2.2	3.9	0.0
U-R	16	20	6.8	6.0	5.9	0.4
U-R	16	20	3.6	16.4	22	0.2
U-R	16	20	0.9	14.8	26.5	0.1
U-R	16	30	6.8	10.0	11.3	0.4
U-R	16	30	3.6	24.3	101.9	0.3
U-R	16	30	0.9	7.6	13.7	0.1
U-R	64	20	8.0	106.2	34.5	2.3
U-R	64	20	4.2	129.8	221.1	0.6
U-R	64	20	1.2	218	286.8	0.5
U-R	64	30	8.0	163.7	134.0	2.6
U-R	64	30	4.2	328.6	1757.1	0.9
U-R	64	30	1.2	253.3	246.3	0.5
UN-R	4	20	5.6	1.1	1.1	0.1
UN-R	4	20	2.9	2.4	5.8	0.0
UN-R	4	20	0.7	0.5	0.9	0.0
UN-R	4	30	5.6	1.8	1.8	0.1
UN-R	4	30	2.9	3.2	22.7	0.1
UN-R	4	30	0.7	0.4	0.9	0.0
UN-R	8	20	6.2	3.7	2.6	0.2
UN-R	8	20	3.4	5.8	11.6	0.1
UN-R	8	20	0.9	2.6	5.8	0.0
UN-R	8	30	6.2	3.8	4.0	0.3
UN-R	8	30	3.4	7.2	29	0.1
UN-R	8	30	0.9	2.4	4.8	0.0
UN-R	16	20	6.8	6.7	6.6	0.4
UN-R	16	20	3.6	20.0	33.6	0.2
UN-R	16	20	0.9	10.6	19.9	0.1
UN-R	16	30	6.8	8.0	9.6	0.4
UN-R	16	30	3.6	18.8	81.8	0.3
UN-R	16	30	0.9	8.5	14.5	0.1
UN-R	64	20	8	108.1	40.5	2.4
UN-R	64	20	4.2	239.4	468.5	0.7
UN-R	64	20	1.2	212.8	358.1	0.5
UN-R	64	30	8	159.4	110.6	2.5
UN-R	64	30	4.2	254.8	946.1	0.8
UN-R	64	30	1.2	314.7	402.1	0.4

Table B.3. Computing time CAC vs Standard



(a) C-AP vs AC-AP **(b)** C-UP vs AC-AP

Fig. B.2. Distribution of the relative performance in low entropy instances

Appendix C

Third Article

C.1. Additional Results

Problem set	I - J - P	u_i	f_i	$[\frac{Z_{90\%}-Z_{10\%}}{Z_{10\%}}]$ $\lvert S\rvert = 500$	$\lvert S\rvert = 1{,}000$	$\lvert S\rvert = 2{,}000$	$\lvert S\rvert = 4{,}000$
IV	16 - 50 -5	5,000	7,500	14.00%	10.47%	6.31%	4.57%
IV	16 - 50 -5	5,000	12,500	14.01%	10.48%	6.32%	4.57%
IV	16 - 50 -5	5,000	17,500	14.01%	10.48%	6.32%	4.57%
IV	16 - 50 -5	5,000	25,000	14.02%	10.49%	6.32%	4.57%
VI	16 - 50 -5	15,000	7,500	14.05%	10.47%	6.35%	4.48%
VI	16 - 50 -5	15,000	12,500	14.05%	10.48%	6.36%	4.49%
VI	16 - 50 -5	15,000	17,500	14.06%	10.48%	6.36%	4.49%
VI	16 - 50 -5	15,000	25,000	14.07%	10.49%	6.36%	4.04%
VII	16 - 50 -5	58,268	7,500	14.05%	10.47%	6.35%	4.48%
VII	16 - 50 - 5	58,268	12,500	14.05%	10.48%	6.36%	4.49%
VII	16 - 50 - 5	58,268	17,500	14.06%	10.48%	6.36%	4.49%
VII	16 - 50 - 5	58,268	25,000	14.07%	10.49%	6.36%	4.49%
IV	16 - 50 - 10	5,000	7,500	12.22%	7.33%	8.94%	6.11%
IV	16 - 50 - 10	5,000	12,500	12.22%	7.34%	8.94%	6.12%
IV	16 - 50 - 10	5,000	17,500	12.23%	7.34%	8.95%	6.12%
IV	16 - 50 - 10	5,000	25,000	12.23%	7.34%	8.95%	6.12%
VI	16 - 50 - 10	15,000	7,500	12.46%	7.33%	8.94%	6.11%
VI	16 - 50 - 10	15,000	12,500	12.47%	7.34%	8.94%	6.12%
VI	16 - 50 - 10	15,000	17,500	12.47%	7.34%	8.95%	6.12%
VI	16 - 50 - 10	15,000	25,000	12.48%	7.34%	8.95%	6.12%
VII	16 - 50 - 10	58,268	7,500	12.46%	7.33%	8.94%	6.11%
VII	16 - 50 - 10	58,268	12,500	12.47%	7.34%	8.94%	6.12%
VII	16 - 50 - 10	58,268	17,500	12.47%	7.34%	8.95%	6.12%
VII	16 - 50 - 10	58,268	25,000	12.48%	7.34%	8.95%	6.12%
VIII	25 - 50 - 5	5,000	7,500	14.08%	10.47%	6.38%	4.48%
VIII	25 - 50 - 5	5,000	12,500	14.09%	10.48%	6.39%	4.48%
VIII	25 - 50 - 5	5,000	17,500	14.09%	10.48%	6.39%	4.48%
VIII	25 - 50 - 5	5,000	25,000	14.10%	10.49%	6.39%	4.48%
IX	25 - 50 - 5	15,000	7,500	14.12%	10.47%	6.42%	4.48%
IX	25 - 50 - 5	15,000	12,500	14.13%	10.48%	6.42%	4.48%
IX	25 - 50 - 5	15,000	17,500	14.14%	10.48%	6.43%	4.48%
IX	25 - 50 - 5	15,000	25,000	14.15%	10.49%	6.43%	4.48%
X	25 - 50 - 5	58,268	7,500	14.12%	10.47%	6.42%	4.48%
X	25 - 50 - 5	58,268	12,500	14.13%	10.48%	6.42%	4.48%
X	25 - 50 - 5	58,268	17,500	14.14%	10.48%	6.43%	4.48%
X	25 - 50 - 5	58,268	25,000	14.15%	10.49%	6.43%	4.48%

Table C.1. Relative differences between $Z_{10\%}$ and $Z_{90\%}$ for optimal solutions with BA-R