Thèse n° 7491

A general framework for the integration of complex choice models into mixed integer optimization

Présentée le 11 septembre 2020

à la Faculté de l'environnement naturel, architectural et construit
Laboratoire transport et mobilité
Programme doctoral en génie civil et environnement

pour l'obtention du grade de Docteur ès Sciences

par

Meritxell PACHECO PANEQUE

Acceptée sur proposition du jury

Prof. M. Andersen, présidente du jury
Prof. M. Bierlaire, Prof. S. Sharif Azadeh, directeurs de thèse
Prof. B. Atasoy, rapporteuse
Prof. B. Gendron, rapporteur
Prof. D. Kuhn, rapporteur

École
polytechnique
fédérale
de Lausanne

2020

"Every portrait that is painted with feeling
is a portrait of the artist, not of the sitter."

— Oscar Wilde, 1890
The picture of Dorian Gray

Acknowledgements

Around five years ago I decided to do a PhD abroad with the objective of pursuing both an international and a research experience. As with other decisions in my life, I was eager and attracted by the idea, but I did not know what it actually entails to do a PhD. Five years later, I can understand why a PhD is often described as an enriching and challenging journey. In this adventure, not only you acquire a profound learning about the topic you are investigating, but you also end up knowing yourself better, both by enjoying the moments of excitement and satisfaction and by learning how to deal with the moments of stress and frustration. In any case, this is not a journey you go on alone, and I want to express my sincere gratitude to everyone that has been part of it.

First, I want to thank Prof. Michel Bierlaire for giving me the opportunity to carry out my PhD under his supervision. I still remember the day I came to EPFL for the interview. That day I had to leave earlier than planned because of heavy delays on the trains going to the airport. I ended up being driven to the airport by a nice professor that I met at the train station who was also about to miss his flight. To my surprise, I received an email from Michel with an offer to start a PhD at TRANSP-OR just when I arrived at the airport, and I forgot about all the stress and inconveniences I experienced that day (if you were wondering, I also caught the flight). From that day on, he has been a great advisor, not only professionally, but also as a person. He has guided me at the academic level, reminded me to believe in myself when I was not doing so and helped me take perspective when I was feeling disoriented. Thank you Michel for your support and for creating a working environment that enhances peer support, collaboration and fun.

The very same day of the interview I met the person who would be my co-advisor in this journey. Prof. Shadi Sharif Azadeh showed me a shortcut to reach the metro station so that I could make it to the train station as soon as possible. From that day on, she has always been kind and understanding, and I have really enjoyed our discussions in all sorts of topics, from research questions to global issues. Thank you Shadi for being always there, even when you were no longer at EPFL.

I would also like to thank the thesis committee for providing me insightful feedback that helped improve this manuscript: Prof. Bilge Atasoy, Prof. Daniel Kuhn and Prof. Bernard Gendron. In particular, I would like to thank Bernard for getting involved in this research, for the discussions on decomposition techniques and for hosting me at

CIRRELT in Montréal. I am also thankful to Prof. Marilyne Andersen for presiding the committee. Despite the inherent pressure of defending a thesis and doing it remotely because of the circumstances, all of them made me feel very comfortable and forget about the stress. Thank you again for everything!

I am also grateful to the Swiss National Science Foundation for supporting this thesis under the project *Incorporating advanced behavioral models in mixed integer linear optimization* (200021-165636).

The working environment has been one of the most important aspects of this experience. I would like to start thanking Mila Bender for all her administrative support and all the conversations about life in general and Morges in particular. I also want to thank Marianne Ruegg, with whom I shortly overlapped but got to know after playing herself in her farewell movie, and Anne Curchod for the IT support. Thanks to Antonin and Flurin for showing me how the end of the journey looked like, Yousef and Matthieu for being always funny and smiling, Marija for transmitting her serenity and for discovering us Serbia together with Darko, Eva for sharing really good organizational and teaching tips, and Iliya for being a wonderful office mate. I am especially grateful to Tom for making me laugh with his *crazy* theories and adventures while traveling, to Stefan for bringing the Swiss culture closer to me and providing always useful feedback, to Riccardo for being always there to listen and to give great advice, and for allowing me to meet Sophia and Elisa, and to Anna for sharing *teenager* hobbies with me, for agreeing on what is *normal* just because we do it like this back home and for giving me the confidence I sometimes do not have (gràcies!).

I am also thankful to the colleagues who joined TRANSP-OR after me and made me feel more senior and go up on the website every time. Thanks to Nicholas and Gael for showing me the amount of things that fit in an office and for helping me to better understand EPFL and Switzerland, Zhengchao for your interest in my research, Yuki for all the chats going to and coming from the MMB classroom, Nikola for being very supportive in the student supervision, Nour for being very generous and caring for others, Claudio for bringing futuristic research to the lab, Rico and Melvin for assessing me at the academic level and Nicola for sharing many interesting things, from cooking recipes to cultural recommendations. I would like to thank Virginie for her supervision during her time at EPFL, for all the deep discussions on research and life and for being a great room mate in conferences, Stefano for being an excellent office mate and for exchanging many ideas related to the framework but also in many different topics, Tim for trying to teach us proper English and for being supportive and funny at the same time, Selin for sharing cool dance videos with me and for being always available to talk about any topic, and Janody for finding the best secret Santa gift and for her involvement in the lab from the day she joined.

The day of the interview I also met Claudia and Leonardo, who started this journey with me, and have also finished it at the same time. Thank you for giving me your vision on each phase of the PhD we were simultaneously going through. I also want to thank the colleagues in LUTS and VITA, and especially Prof. Nikolas Geroliminis and Prof. Alexandre Alahi, for their feedback during the internal seminars.

Teaching has been a very nice experience too, and I have learned from all the students I had the opportunity to supervise. Thank you Jonathan, Martí, Thibaut, Gabriel, Yassine, Oliver, Younes, Tatiana, Benjamin and especially Jérémy, to whom I want to thank for his dedication and help with research ideas for this thesis.

During these 5 years, I have coincided with many people from many different countries and with different backgrounds in conferences, parties and other events. This has taught me the importance of getting together and that, after all, we are all humans. I would also like to thank the friends that, regardless of the distance, with a simple message asking how it was going or a short visit have made me feel closer to home.

Quiero agradecer a mi familia por el apoyo que he recibido y por estar siempre tan unidos. En especial agradezco a mis padres la libertad que me han dado y la confianza que han depositado en mí en todo momento, dejándome tomar mis propias decisiones y estando ahí para ayudarme con lo que hiciera falta. También quiero darles las gracias a mis hermanos Gustavo y Gisela por hacer que todo fuera menos aburrido y hacerme un poco más responsable y generosa.

Y por último, gracias a ti, Aitor, por embarcarte también en esta aventura y haber hecho que los buenos momentos fueran aún mejores y los malos fueran mucho más llevaderos. Gracias por comprenderme y por ayudarme en todo, especialmente al final de esta etapa.

Morges, 17th June 2020 Meritxell Pacheco Paneque

Résumé

Le but de cette thèse est de développer une méthodologie générale afin d'incorporer une représentation de la demande désagrégée dans des problèmes d'optimisation orientés vers l'offre qui permette de capter l'interaction entre le comportement des individus et les décisions à optimiser. Pour cela, nous présentons un cadre de modélisation pour intégrer des modèles de choix discret (MCD) dans les problèmes linéaires à variables mixtes (PLM), et nous montrons qu'il est à la fois flexible et opérationnel sur des cas réalistes. En particulier, nous développons des algorithmes pour augmenter la tractabilité de ce cadre, et nous illustrons son applicabilité avec deux problèmes d'optimisation présents dans de nombreux contextes.

Les fonctions de demande résultants des MCD sont fortement non-linéaires et non-convexes, et elles ne sont pas toujours données dans une forme explicite. Dans cette thèse, nous évitons l'usage de ces fonctions en précisant la structure des préférences du MCD directement en termes des équations structurales associées (fonctions d'utilité). Le traitement de la nature probabiliste de ces équations se fait par simulation, avec des tirages de la distribution du composant aléatoire associé. Cela produit un ensemble de restrictions linéaires à variables mixtes qui peut être intégrée dans une formulation PLM quelconque. La seule exigence est que les décisions à optimiser qui sont aussi des variables explicatives du MCD, et donc captent les interactions, apparaissent linéairement dans les équations structurales.

La nature désagrégée des MCD, ainsi que la linéarisation basée sur la simulation associée, comporte une complexité de calcul élevée. Inspirés par la structure décomposable du cadre sur les deux dimensions sur lesquels il est construit, les individus et les simulations de tirages, nous caractérisons une approche de décomposition lagrangienne qui permet résoudre des cas de plus grande taille, au moins de façon approximative. En effet, les tests réalisés montrent que des solutions presque optimales sont obtenues avec un temps de calcul fortement réduit (en utilisant seulement 10% du temps de calcul utilisé par la méthode exacte).

Ce cadre est assez général pour s'adapter à une grande variété de problèmes d'optimisation habituels. Son principal point fort est qu'il n'est pas nécessaire d'adapter le MCD à la formulation, ce qui permet de l'emprunter directement de la littérature. En particulier, on ne se limite pas à des MCD avec des hypothèses simplistes, comme par exemple le modèle logit, et nous pouvons intégrer des MCD plus avancés, notamment des modèles logit mixtes. Dans cette thèse, nous considérons et résolvons deux problèmes afin d'illustrer la versatilité du cadre, à savoir la maximisation du profit de l'opérateur et la conception d'un système de transport orienté vers les voyageurs. Pour le premier nous considérons un opérateur qui fournit des services à un marché avec l'objectif de maximiser son profit. Pour le second nous formulons la tarification et la conception d'un système de transport en maximisant une mesure de bien-être social. L'élément quantitatif clé de l'analyse du bien-être dans le contexte des MCD, l'utilité maximale espérée, est aisément exploitable dans le cadre proposé. Cela représente un avantage significatif car les formulations non-linéaires complexes qui résultent de l'intégration de cet élément ne sont plus nécessaires.

En résumé, cette thèse fait des contributions importantes à l'intégration des MCD dans les formulations PLM, et montre leur applicabilité sur des vrais problèmes d'optimisation. Les modèles et algorithmes proposés mettent en lumière les avantages d'inclure le comportement individuel dans les décisions opérationnelles de toute industrie montrant une forte interaction entre les décisions concernant l'offre et la demande.

Mots-clés: Modélisation mathématique du comportement, demande désagrégée, modèles de choix discret, optimisation combinatoire, problèmes linéaires à variables mixtes, décomposition lagrangienne, maximisation du profit orientée à l'opérateur, planification de transport orientée à l'utilisateur, analyse du bien-être.

Abstract

The objective of this thesis is to develop a general methodology to incorporate a disaggregate demand representation in supply-oriented optimization problems that allows to capture the interplay between the behavior of individuals and the decisions to be optimized. To this end, we propose a modeling framework for the integration of discrete choice models (DCM) in mixed-integer linear problems (MILP), and we show that it is both flexible and operational on realistic instances. In particular, we develop algorithms to enhance the tractability of the framework, and we illustrate its applicability with two relevant optimization problems that arise in a great deal of contexts.

The demand functions generated from DCM are highly non-linear and non-convex, and are not always available in closed form. In this thesis, we avoid the use of such functions by specifying the preference structure of DCM directly in terms of the related structural equations (utility functions). We rely on simulation in order to handle the probabilistic nature of these equations by drawing from the distribution of the associated random component. This yields a mixed-integer linear set of constraints that can be embedded in any MILP formulation. The only requirement is that the decisions to be optimized that are also explanatory variables of the DCM, and therefore capture the interactions, appear linearly in the structural equations.

The disaggregate nature of DCM, together with the associated simulation-based linearization, comes with a high computational complexity. Motivated by the decomposable structure of the framework along the two dimensions it is built on, the individuals and the simulation draws, we characterize a Lagrangian decomposition scheme that enables to solve larger instances, at least approximatively. Indeed, the performed tests show that near-optimal solutions are obtained in a much reduced computational time (by running only 10% of the computational time used by the exact method).

The framework is sufficiently general to accommodate a wide variety of relevant optimization problems. The main strength is that the DCM does not need to be tailored to the formulation, i.e., it can be taken as such from the literature. In particular, it does not have to be a DCM that relies on simplistic assumptions, such as the logit model, and more advanced DCM such as mixtures of logit models can be integrated. In this thesis, we consider and solve two problems in order to illustrate the versatility of the framework, namely operator-centric profit maximization and traveler-centric design of a transportation system. The former assumes an operator that offers services to a market with the aim of maximizing its profit. The latter formulates the pricing and design of a transportation system such that a measure of welfare is maximized. The key quantitative element of welfare analysis in the context of DCM, the expected maximum utility, is readily available in the framework. This represents a significant advantage because it allows not to deal with the complex non-linear formulations that result from the integration of this quantity as provided by the discrete choice theory.

In summary, this thesis makes relevant contributions on the integration of DCM in MILP, and shows their applicability by relying on real-world optimization problems. The proposed models and algorithms shed some light on the benefits of incorporating individual behavior in operational decisions for any industry with close interactions between the demand and the supply.

Keywords: Mathematical modeling of behavior, disaggregate demand, discrete choice models, combinatorial optimization, mixed-integer linear problems, Lagrangian decomposition, operator-centric profit maximization, user-centric transportation planning, welfare analysis.

Resumen

El objetivo de esta tesis es el desarrollo de una metodología general con el fin de incorporar una representación de la demanda de manera desagregada en problemas de optimización orientados a la oferta que permita captar la interacción entre el comportamiento de los individuos y las decisiones a optimizar. Para ello se propone un marco de modelización para la integración de modelos de elección discreta (MED) en programación lineal entera-mixta (PLEM) y se comprueba que dicha estructura es flexible y operacional en ejemplos realistas. En concreto, se desarrollan algoritmos que aumentan la tratabilidad de este marco y se ilustra su aplicabilidad con dos problemas de optimización que surgen en múltiples ámbitos.

Las funciones de demanda que se derivan de los MED son altamente no lineales y no convexas, y no siempre están disponibles en forma cerrada. En esta tesis se evita el uso de dichas funciones al especificar la estructura de preferencias relativa a los MED directamente en términos de las ecuaciones estructurales correspondientes (funciones de utilidad). La naturaleza estocástica de dichas ecuaciones es abordada con la generación de realizaciones aleatorias de la distribución de la componente aleatoria asociada. Esto da lugar a un conjunto de restricciones lineales con variables mixtas que puede incluirse en cualquier formulación PLEM. El único requisito es que las decisiones a optimizar que también sean variables explicativas del MED, y que por lo tanto captan las interacciones, aparezcan de manera lineal en las ecuaciones estructurales.

La naturaleza desagregada de los MED, junto con la correspondiente linearización basada en simulación, comporta una complejidad computacional elevada. Motivada por la estructura descomponible del marco de modelización en las dos dimensiones sobre las que se construye, los individuos y las realizaciones, se desarrolla una descomposición lagrangiana que permite resolver problemas de gran tamaño de manera aproximada. En efecto, en los tests realizados se obtienen soluciones prácticamente óptimas en un tiempo de computación mucho más reducido (únicamente un 10% del tiempo de computación empleado por el método exacto).

Este marco es lo suficientemente general como para albergar una amplia variedad de problemas de optimización. La principal ventaja es que el MED no se tiene que adaptar a la formulación, lo que permite implementarlo tal y como aparece en la bibliografía. En particular, no es necesario que sea un MED basado en hipótesis simplistas como por ejemplo el modelo logit, ya que otros MED más complejos, por ejemplo el modelo logit mixto, pueden ser directamente integrados. En esta tesis se consideran y se resuelven dos problemas con el objetivo de ilustrar la versatilidad del marco, concretamente la maximización del beneficio por parte de un operador y la planificación de un sistema de transporte orientado a los pasajeros. Para el primero se considera un operador que ofrece servicios a un mercado de manera que se maximiza el beneficio generado por su venta. Para el segundo se formula el diseño y la tarificación de un sistema de transporte con el objetivo de maximizar una medida de bienestar social. El elemento cuantitativo clave para el análisis del bienestar en el contexto de los MED, la utilidad máxima esperada, se puede obtener de manera inmediata gracias al marco propuesto. Esto representa una ventaja significativa porque las formulaciones no lineales que resultan de la integración de dicho elemento ya no son necesarias.

En resumen, esta tesis hace contribuciones relevantes en relación a la integración de los MED en formulaciones de PLEM, y demuestra su aplicabilidad en problemas de optimización que surgen en el mundo real. Los modelos y algoritmos propuestos en esta tesis arrojan luz sobre las ventajas de incluir el comportamiento a nivel individual en las decisiones operacionales en cualquier sector con fuertes interacciones entre la oferta y la demanda.

Keywords: modelización matemática del comportamiento, demanda a nivel desagregado, modelos de elección discreta, programación lineal entera-mixta, descomposición lagrangiana, maximización del beneficio con respecto a un operador, planificación de transporte orientada al pasajero, análisis del bienestar.

Contents

Acknowledgements **i**

Résumé **v**

Abstract **vii**

Resumen **ix**

List of Figures **xv**

List of Tables **xvii**

1 Introduction **1**
1.1 Context . . . 1
1.2 State of the art and motivation . . . 3
1.3 Research objectives and scientific contributions . . . 10
1.4 Structure of the thesis . . . 12

2 Choice-based optimization framework **15**
2.1 Introduction . . . 15
2.2 Modeling framework . . . 16
2.2.1 The discrete choice model . . . 17
2.2.2 Simulation-based linearization . . . 21
2.2.3 Capacities as decision variables . . . 26
2.2.4 Size of the mixed-integer linear formulation . . . 27
2.2.5 The optimization model . . . 28
2.3 Profit maximization problem . . . 29
2.3.1 Continuous price . . . 30
2.3.2 Discrete price . . . 31
2.3.3 Expected profit . . . 32
2.4 Case study . . . 33
2.4.1 Price calibration . . . 36
2.4.2 Population segmentation . . . 40
2.4.3 Impact of the arrival of individuals . . . 44
2.4.4 Profit maximization through capacity allocation . . . 46

2.5 Concluding remarks . . . 47

3 A Lagrangian decomposition scheme for the choice-based optimization framework 49

3.1 Introduction . . . 49
3.2 Choice-based revenue maximization problem . . . 51
3.3 Literature review . . . 55
3.3.1 Complicating variables . . . 55
3.3.2 Complicating constraints . . . 57
3.3.3 Discussion . . . 60
3.4 Initial attempts on the Lagrangian relaxation scheme . . . 63
3.4.1 The customer and the operator subproblems . . . 63
3.4.2 Lagrangian decomposition on the price variables . . . 67
3.5 Lagrangian decomposition scheme . . . 71
3.5.1 Decomposition by subsets of draws . . . 71
3.5.2 Feasible solutions . . . 73
3.5.3 Subgradient method . . . 76
3.6 Case study . . . 79
3.6.1 Comparison with optimal solutions . . . 80
3.6.2 Trade-off between R, N and S . . . 82
3.7 Concluding remarks . . . 86

4 Welfare-maximizing design of a transportation system 89

4.1 Introduction . . . 89
4.2 Literature review . . . 91
4.3 Methodology . . . 95
4.3.1 Social welfare . . . 97
4.3.2 Budgetary constraint . . . 99
4.3.3 Demand-dependent travel time . . . 101
4.3.4 Model-based algorithmic approach . . . 103
4.4 Introduction of a congestion toll in an urban context . . . 104
4.4.1 ICLV model . . . 106
4.4.2 Assumptions about the demand . . . 109
4.4.3 Assumptions about the supply . . . 111
4.4.4 Numerical results . . . 112
4.5 Evaluation of a dedicated bus lanes policy . . . 116
4.5.1 Nested logit model . . . 117
4.5.2 Optimization problem . . . 120
4.5.3 Linear formulation . . . 122
4.5.4 Numerical results . . . 125
4.6 Concluding remarks . . . 129

5 Conclusion **131**
5.1 Main findings and implications 131
5.2 Future research directions 133

A Notations of the choice-based optimization framework **137**

B Convergence of the choice-based optimization framework **139**

Bibliography **147**

Curriculum Vitae **161**

List of Figures

2.1 Expected revenue obtained from group 1, group 2 and total as a function of the price of alternative i (p_i) 20
2.2 Boxplot of the optimal prices for the uncapacitated case for 5 replications (independent generation of draws) 37
2.3 Boxplot of the expected demand for the uncapacitated case for 5 replications (independent generation of draws) 37
2.4 Boxplot of the optimal prices for the capacitated case for 5 replications (independent generation of draws) 38
2.5 Boxplot of the expected demand for the capacitated case for 5 replications (independent generation of draws) 39
2.6 Optimal prices for the user groups described in Section 2.4.2 for for 5 replications (independent generation of draws) 44
2.7 Expected demand for the user groups described in Section 2.4.2 for 5 replications (independent generation of draws) 44
2.8 Histogram of the expected optimal revenue in situation 2 (discount rate of 30%) for 100 priority lists and $R = 50$ (individual level) 45
2.9 Histogram of the expected optimal revenue in situation 2 (discount rate of 30%) for 100 priority lists and $R = 50$ (group level) 45
2.10 Histogram of the expected optimal revenue in situation 2 (discount rate of 30%) for 100 priority lists and $R = 100$ (group level) 46

3.1 Graphical representation of p^*_{inr} 70
3.2 Evolution of the lower and upper bounds throughout the iterations of the Lagrangian decomposition scheme for $N = 50$, $R \in \{100, 250, 500\}$ and $S = 5$ 82
3.3 Duality gap (gap_{dual}) and iteration at which the best lower bound is achieved (in brackets) for the 3 considered instances (1-3) for $R \in \{100, 250, 500, 1000, 2500, 5000\}$ 84
3.4 Duality gap (gap_{dual}) and iteration at which the best lower bound is achieved (in brackets) for the 3 considered instances (1-3) for $N \in \{50, 100, 150, 197\}$ 85

3.5 Duality gap (gap_{dual}) and iteration at which the best lower bound is achieved (in brackets) for the 3 considered instances (1-3) for $S \in \{1, 2, 3, 4, 5, 10\}$ . . . 86

4.1 Geographical division of the Lausanne-Morges region into subregions. Source: Guillaume-Gentil et al. (2015) . . . 105
4.2 Distribution of ν_n in the definition of the waiting time (4.37) . . . 111
4.3 Difference in social welfare [CHF] with respect to the base case for the 3 considered replications (independent generation of draws, labeled as 1-3) for $R \in \{10, 25, 50, 100, 250, 500\}$ in the non-congested situation . . . 113
4.4 Difference in social welfare [CHF] with respect to the base case for the 3 considered replications (independent generation of draws, labeled as 1-3) for $R \in \{10, 25, 50, 100, 250, 500\}$ in the partially congested situation . . . 114
4.5 Difference in social welfare [CHF] with respect to the base case for the 3 considered replications (independent generation of draws, labeled as 1-3) for $R \in \{10, 25, 50, 100, 250, 500\}$ in the highly congested situation . . . 114
4.6 Decision tree associated with each individual (nesting structure) . . . 118
4.7 Linear formulation of the dedicated bus lanes policy (Basso and Silva, 2014) . . . 125
4.8 Difference in consumer surplus [\$/day-km] between DL and DL+CON+SUB for the 3 considered replications (independent generation of draws, labeled as 1-3) for $R \in \{100, 250, 500, 1000\}$ (ref. represents the value reported by Basso and Silva (2014)) . . . 127
4.9 Average variation in consumer surplus [\$/pax] between DL and DL+CON+SUB for the 3 considered replications (independent generation of draws) for $R = 1000$ . . . 128

List of Tables

2.1 Specification table of the mixtures of logit model in Ibeas et al. (2014) . . 34
2.2 Computational results of the revenue maximization problem for the uncapacitated cases for 5 replications (independent generation of draws) 37
2.3 Computational results of the revenue maximization problem for the capacitated case for 5 replications (independent generation of draws) 38
2.4 Relative differences of the expected optimal revenues obtained by solving the MILP model with respect to the expected optimal revenues obtained by evaluating the optimal prices with $R^* = 10^6$ for 5 replications (independent generation of draws) 39
2.5 Average revenue and range of obtained results for 5 replications (independent generation of draws) for different subsidies in situation 1 for $R = 50$ (R denotes resident and NR denotes non resident) 41
2.6 Average revenue and range of obtained results for 5 replications (independent generation of draws) for different subsidies in situation 2 for $R = 50$ (R denotes resident and NR denotes non resident) 41
2.7 Definition of user groups 42
2.8 Computational results for the user groups defined in Table 2.7 for 5 replications (independent generation of draws) 43
2.9 Results for the profit maximization with capacities as decision variables for 5 replications (independent generation of draws) for $R = 25$ (in brackets the number of replications for which the associated capacity level was obtained) 47

3.1 Comparison of the performance of the exact method and the Lagrangian decomposition scheme for $N = 50$, $R \in \{100, 250, 500\}$ and $S = 5$ 81
3.2 Comparison of the prices and the expected demand of the optimal solution (Ex.) and the best feasible solution (Dec.) found in the Lagrangian decomposition scheme for $N = 50$, $R \in \{100, 250, 500\}$ and $S = 5$ 81
3.3 Computational performance of the Lagrangian decomposition scheme for the 3 considered instances defined by $N = 50$, $S = 2$ and $R \in \{100, 250, 500, 1000, 2500, 5000\}$ 83

3.4 Computational performance of the Lagrangian decomposition scheme for the 3 considered instances defined by $R = 500$, $S = 2$ and $N \in \{50, 100, 150, 197\}$ . . . 84
3.5 Computational performance of the Lagrangian decomposition scheme for the 3 considered instances defined by $N = 50$, $R = 500$ and $S \in \{1, 2, 3, 4, 5, 10\}$ . . . 85

4.1 Specification table of the utilities of the ICLV model (Atasoy et al., 2013) 107
4.2 Specification table of the structural equations for the attitudes of the ICLV model (Atasoy et al., 2013) . . . 108
4.3 Average computational time (min) of Algorithm 4.1 for 3 replications (independent generation of draws) for the optimized case . . . 115
4.4 Design variables (p_{PMM}, p_{PT} and f) for the 3 considered replications (independent generation of draws, labeled as 1-3) for $R = 500$ (multiple lines per replication indicate the presence of a cycle, i.e., the algorithm terminated because of a cycle and two solutions were reported) . . . 115
4.5 Modal split (%) and travel time for PMM [min] for the base case and the optimized case for the 3 considered replications (independent generation of draws, labeled as 1-3) for $R = 500$ (multiple lines per instance indicate the presence of a cycle, i.e., two solutions were reported) . . . 116
4.6 Estimates of the parameters of the nested logit model defined by (4.39)–(4.40) (Basso and Silva, 2014) . . . 119
4.7 Average computational time (min) of Algorithm 4.1 for 3 replications (independent generation of draws) for the problem in Model 4.7 and the two evaluated policies . . . 127
4.8 Comparison of the range of flows [veh/h] obtained with 3 replications (independent generation of draws) with the results obtained in Basso and Silva (2014) for $R = 1000$ . . . 128
4.9 Comparison of the range of values of the design variables obtained with 3 replications (independent generation of draws) with the results obtained in Basso and Silva (2014) for $R = 1000$ . . . 128

A.1 Main notations used in the choice-based profit maximization problem (sets, parameters, variables and aggregated quantities). For the sake of simplicity, we redefine the set of alternatives $\mathcal{C}$ as the set of duplicates associated with each capacity level q, and we therefore denote c_{iq}, f_{iq} and v_{iq} simply by c_i, f_i and v_i, respectively. . . . 138

1

Introduction

1.1 Context

A mismatch between supply and demand is the imbalance between the amount of supplies of a product or service with the corresponding willingness or need in the market. It represents a concern that affects a great deal of contexts, such as transportation, supply chain, health and manufacturing, to name a few. This asymmetry results in multiple consequences, which include reduced profitability, a decrease in consumer confidence and spillover effects. Revenue management (RM) and the analysis of welfare measures represent two relevant examples that allow us to illustrate the extent of this subject.

RM has been widely investigated by academics and practitioners in the airline industry (McGill and van Ryzin, 1999). The research on airline RM revolves around pricing, seat inventory control, overbooking and demand forecasting. The latter strongly influences the other aspects, as demand forecasting has an effect on the booking limits, which determine the profits, and overbooking calculations depend on predictions on passenger cancellations and no-shows. The success of airline RM has stimulated the development of RM systems in other transportation sectors (e.g., passenger railways) and other areas of the service sector (e.g., hospitality).

Welfare measures are defined in different settings to evaluate the performance of a policy that wants to be put into practice by a public authority. Some examples can be found in the labor market concerning unemployment insurance (Mukoyama, 2013) and in transportation to correct the negative externalities associated with road traffic (Parry and

Bento, 2001). A detailed demand representation enables to determine if the policy is regressive, i.e., the welfare is distributed more unequally after its introduction than before, and to identify which segments in the population are the most adversely impacted.

These two application areas illustrate the necessity of an appropriate demand representation and the importance of explicitly allowing for the interplay between the individuals (demand) and the design and planning decisions to be made by the operator (supply), such as the price of a flight ticket or the level of subsidy on a public transportation mode. Aggregate representations of the demand are commonly used in the optimization models that formulate operators' decisions (Bierlaire and Lurkin, 2017). Nevertheless, as the demand is the result of the decisions performed by individual actors, an aggregate modeling is not able to capture the causal mechanisms that generate the demand. Furthermore, in order to characterize the reaction of demand to changes in the operators' decisions, such decisions need to be explanatory variables of the demand model. If they are exogenous to the demand model, their impact on the individuals' behavior is not properly represented, which leads to an unrealistic description of the system. Indeed, if the system is designed for the average consumer, individuals deviating from the mean will not be satisfied, either because what they like is overpriced, or what they can afford does not provide them the requested level of service.

Discrete choice models (DCM) are the state-of-the-art of demand modeling at the disaggregate level when the outcome of the decision-making process of individuals is discrete. Rooted in the theoretical foundations of microeconomics, DCM allow to capture the causality between the explanatory variables (attributes of the alternatives and socioeconomic characteristics of the individuals) and the choice itself thanks to the concept of random utility. The output of these models are the probabilities associated with each individual to choose each alternative, which enables to represent the heterogeneity of tastes and preferences in high detail. These probability expressions are non-linear and non-convex in the explanatory variables, and might not even have a closed-form.

On the other hand, the optimization models that describe the supply-related decisions typically require linearity or convexity of the involved mathematical functions in order to ensure tractability of the resulting formulation. In fact, in convex optimization problems, and consequently in linear optimization problems, a local optimum is also a global optimum. As optimization algorithms are designed to identify local optima, this property guarantees that a global optimum will be found by the algorithm.

Thus, embedding a disaggregate demand representation given by a DCM in an optimization model while explicitly allowing for the supply-demand interactions is a difficult task, not only from the modeling point of view, but also with respect to the design of solution methodologies. Section 1.2 provides an overview of the literature related to such integration and motivates the work undertaken here.

1.2 State of the art and motivation

As a consequence of the difficulty in generating demand data, demand has been typically assumed as a given input in many problem instances addressed in the operations research literature (Bierlaire and Lurkin, 2017). For example, in the context of facility location, the complexity of the formulations modeling the decisions on spatial resource allocation has historically limited the research to deterministic problems, where all input parameters (including the demand) are considered as known (Laporte et al., 2015). As in reality demand might broadly fluctuate, it is quite common to use aggregate statistical methods, such as time series analysis, to predict the demand data. In facility location, researchers typically assume that the demand follows a probability distribution or changes its patterns under different hypothetical scenarios. Nevertheless, as pointed out in Section 1.1, an aggregate demand modeling cannot represent the underlying causal mechanisms that generate the demand.

Since the popularization of the logit model (McFadden, 1974), and thanks to the increasingly availability of abundant and individual-based datasets, DCM based on the random utility principle have become the most advanced and operational disaggregate demand models. The key advantage of the logit model relies on the simplicity of its closed-form probability expression. However, the associated assumptions might give rise to unrealistic substitution patterns across alternatives. More complex models have been proposed to relax such assumptions. Nested logit and cross nested logit models allow to account for correlations between alternatives by defining nests that group alternatives with common features. Like the logit model, they have a closed-form expression. Furthermore, it has been shown that the choice probabilities of any (additive) random utility model might be approximated by a cross-nested logit model (Fosgerau et al., 2013). Mixtures of logit models, also known as mixed logit models, provide a highly flexible framework that can also approximate any DCM (McFadden and Train, 2000). Their attractiveness lies on the fact that they overcome the main limitations of standard logit models by allowing for random taste variation, unrestricted substitution patterns and correlation in unobserved factors over time. In addition to capturing the variation of preferences across individuals (inter-consumer heterogeneity), more advanced versions of mixed logit models defined at the individual level (Hess and Rose, 2009, Becker et al., 2018) also enable to incorporate the variation of preferences for the same individual across different choice situations (intra-consumer heterogeneity).

Hence, more complex DCM are able to better forecast the behavior of individuals. Nevertheless, the focus on behavioral realism results in sophisticated mathematical formulations that are difficult to embed in classical operations research models. For instance, the choice probabilities associated with mixed logit models do not have a closed-form since they involve integrals of standard logit probabilities over a density of parameters. Moreover, as the choice probabilities are non-linear and non-convex in the explanatory variables (which include the optimization variables that have an impact on the individuals' behavior, and therefore capture the supply-demand interactions), so are the aggregate demand indicators derived thereof, such as the expected revenue or the welfare measure. As a consequence, obtaining a convex formulation with advanced DCM is unattainable, which significantly complicates their use in exact methods of optimization.

Mixed-integer linear problems (MILP) are optimization problems that involve only linear functions and accommodate both integer and continuous variables. The fact that these problems naturally arise in many contexts that simultaneously face discrete and continuous decisions (e.g., facility location, RM, transportation-related problems) has led to an increased interest and a significant development in the last decades (Vielma, 2015). The main reasons for the success of MILP models are the modeling flexibility that these formulations provide and the availability of linear programming (LP) solvers. Indeed, note that MILP are not convex problems, not even linear. What is crucial here is that its relaxation is convex or, better, linear. This property is needed to design solution algorithms based on branch-and-bound.

There exists a great variety of formulation strategies and solution techniques in MILP. There are often alternative MILP formulations associated with an optimization problem. Some models might be smaller in terms of the number of variables and constraints, but might be more difficult to solve than larger models, and can make a difference in whether or not MILP formulations can be solved quickly enough to be practically useful (Smith and Taskin, 2008). With respect to the solution methods, commercial MILP solvers, which employ a combination of branch-and-bound and cutting-plane techniques, are available to practitioners and analysts to find the global optima of MILP. They can therefore be used as a *black box*, which allows the user to focus on the modeling instead of the development of a solution algorithm. Nevertheless, if the problem at hand has a special structure that can be exploited, the solution algorithm can be adapted in order to increase the performance of the solver.

As MILP are NP-hard, many problem instances might be intractable, and heuristic methods must be used instead. MILP heuristics aim at finding a feasible (and hopefully good) solution of MILP. The availability of very effective general-purpose heuristics for MILP formulations has represented a fundamental improvement (Fischetti and Lodi, 2010). Furthermore, a large number of MILP can be viewed as potentially *easy* problems to solve that are complicated by a certain set of variables and/or constraints (Fisher, 1981a). Lagrangian relaxation (Geoffrion, 1974) and Benders decomposition (Benders, 1962) are the two classical strategies that exploit the decomposable structure of the problem in order to generate blocks that can be independently addressed and are less complex to handle. Decomposition techniques provide comprehensive guidelines on the separability of MILP formulations, and represent a flexible scheme that can be combined with additional heuristic approaches to increase the efficiency of the solution algorithm.

Notwithstanding the discussed advantages and existing developments associated with DCM to represent the demand and with MILP to describe the supply-related decisions, the distinct nature of both modeling approaches makes their integration rather challenging. Research in this direction is receiving increased attention in different fields where the demand representation plays a key role in the decisions that want to be optimized. We review some relevant works on the integration of DCM in optimization models, with a special emphasis on MILP, in the context of facility location, RM and transportation-related problems.

In the last years, there has been a growing body of literature on facility location problems relying on DCM (mainly the logit model) to represent the behavior of customers. The inclusion of the choice probabilities in the problem of locating p new facilities in a competitive market such that the captured demand is maximized, known as the maximum capture problem, has been given a significant consideration. In Benati (1999), the resulting optimization model is reformulated as a p-median problem and solved quickly by Lagrangian relaxation and branch-and-bound. Later, in Benati and Hansen (2002), three solution methods are developed (one exploiting the concavity of the objective function and the other two reformulating the problem as an MILP model), and it is shown that only moderate size problems can be solved up to optimality.

The maximum capture problem described in Haase (2009) considers homogeneous customers within the blocks that discretize the area of interest, and obtains the choice probabilities of a reduced choice set by linearly modeling the property of constant substitution patterns associated with the logit model. The author also allows for flexible substitution patterns by simulating multiple individuals within each block. For each individual, the associated utility values are obtained by generating the stochastic component at random. The same idea is exploited in Müller et al. (2009) and Haase and Müller (2013) in school location. The former describes a two-step approach to minimize the location and transportation costs with respect to students choosing the school with the highest utility (given by a mixture of logit models). The first step (quadratic constrained problem) allocates students for each scenario (defined as a combination of open and closed schools) based on utility and capacity, and in the second step (combinatorial problem), the scenario minimizing the total costs is chosen. The latter formulates an integer formulation for the same problem that can be solved optimally (or at least close to optimality) within a few minutes with commercial solvers.

Similar approaches mostly focus on developing MILP models that linearize the resulting objective function (e.g., Aros-Vera et al., 2013 in park and ride facility location and Zhang et al., 2012 in health care facility location). Haase and Müller (2014) compare different MILP reformulations, and the results show that the one proposed by Haase (2009) for solving large problems using a commercial solver is the most promising one. In Freire et al. (2016), this benchmark is extended and the computational study shows inconclusive results, as different methods performed dissimilarly depending on the considered dataset. Outer approximation, an algorithm used to solve mixed-integer non-linear problems (MINLP) based on decomposition principles (Duran and Grossmann, 1986), has been recently considered in Ljubić and Moreno (2018). They define a MILP formulation based on outer approximation cuts and provide a branch-and-cut procedure that outperforms the best performing strategies according to Freire et al. (2016).

Despite its relevance, demand has typically been assumed to be isolated from its market environment in airline RM systems (van Ryzin, 2005), and the lack of information about customers' preferences and the complexity of the resulting mathematical formulations has made disaggregate forecasting extremely difficult and infrequently used in practice (Talluri and Van Ryzin, 2006). However, customer-behavior-oriented models of demand represent a promising approach for RM, especially DCM (Shen and Su, 2007). Talluri and Van Ryzin (2004) provide an exact analysis of the impact of choice behavior in RM by explicitly modeling consumers' behavior with a general DCM where the probabilities of purchasing each fare product depend on the set of available fare products, and show that significant improvements in the revenue can be achieved with respect to the traditional methods.

The models integrating DCM in RM optimization models are known as choice-based RM models, and have gained popularity over the last years. Such models generally aim at maximizing both revenue and customer satisfaction by deciding about pricing while controlling for product availability. They were first introduced by Andersson (1998), where a logit model is assumed to compute the probability of a passenger that was rejected at one flight-class combination to request a seat at another flight-class, called the recapture rate (or buy-up rate). An example is considered to show that the revenue increases when recapture and buy-up are implemented. Ratliff et al. (2008) extends this work by involving the solution to the problem solved by Andersson (1998) in a heuristic approach, that is, given historical bookings and RM controls, the goal is to find the unconstrained first-choice demand, spill and recapture, which can be considered as the inverse problem of the RM optimization. The resulting methodology can be readily integrated into existing demand forecasting systems and should reduce forecast errors.

Schön (2007) develops a theoretical market-oriented model for airline network service design that integrates flight schedule design, fleet assignment and pricing decisions. For each customer segment, the demand is explicitly defined as a function of the price that needs to satisfy certain properties, which are met by the logit model and the nested logit model, as later specified in Schön (2008). Under suitable assumptions, the resulting formulation is a mixed-binary maximization problem with concave objective function and linear constraints that can be solved with standard techniques. Atasoy et al. (2014) address a similar problem by also including spill and recapture effects based on the logit model. The resulting formulation is a mixed-integer non-convex problem for which solvers can provide good quality feasible solutions for instances with moderate size. More recently, Korfmann (2018) develops a single-leg MILP model with flexible demand substitution patterns between fare classes, where the demand is represented by the individual utility values (using Monte Carlo simulation), and the objective is the optimal allocation of bookings to the offered fare classes.

In several transportation networks, travelers may modify their travel arrangements (e.g., departure time, route) depending on the level of service of the network and/or the price associated with the available alternatives. Traffic assignment is an essential element in transportation planning, as it allows, among others, to assess the deficiencies in the existing transportation system, to test alternative system proposals and to provide real-time guidance (Patriksson, 2015). Recent works in dynamic traffic assignment (Pel et al., 2009, Qian and Zhang, 2013) investigate hybrid route choice models (usually relying on the logit model) where all travelers have a pre-trip route but consider real-time traffic conditions in seeking new routes. As opposed to classical approaches, where route choices are modeled by assuming that travelers simply choose the cheapest or shortest route presented to them, a better description of the system (e.g., congestion, queue spillovers) is achieved with hybrid route choice models. In spite of its closed form, the logit model has received particular criticism for its independently and identically distributed error term assumption (Sheffi, 1984), which has motivated researchers to relax such assumptions by proposing variants of the standard logit model (e.g., Chen et al., 2012) or alternative approaches that assume the knowledge of marginal distributions of the error terms but do not assume that the random error terms are independently or identically distributed (e.g., Damla Ahipaşaoğlu et al., 2016).

The toll setting problem was introduced by Labbé et al. (1998), and consists of a bilevel model where the authority (upper level) wants to maximize its revenues from a taxation scheme on a transportation network at the same time that the users (lower level) minimize their generalized travel costs (i.e., the tolls and the travel costs) while allowing for those tax levels. The model is formulated as bilevel program with bilinear objectives at both levels of decision and network constraints at the lower level. It can be efficiently reformulated as a MILP with a small number of binary variables, large instances of which can be solved within reasonable time. Nevertheless, as mentioned by the authors, the considered deterministic representation of user behavior is too simplistic, since it assumes no dispersion of traffic along the routes and the value of time is uniform throughout the population.

The stochastic version of the toll setting problem that assigns users to paths according to a DCM has received less attention in the literature than its deterministic counterpart. In Gilbert et al. (2014a), a logit route choice model is used to account for users' awareness of the network conditions. The optimization problem is non-linear and non-convex, and may have several local optima. An exhaustive numerical study of this problem is carried out in Gilbert et al. (2015), where it is concluded that the problem can be solved for a near-optimal solution by a combination of mixed-integer approximations and local ascent methods. In Gilbert et al. (2014b), the use of a mixture of logit models (price sensitivity distributed across users) makes this approach numerically challenging as no closed-form solution is available for the assignment of users to paths. Hence, approximation schemes that provide starting points from which a local search converges to a near-optimal solution are implemented.

The inclusion of DCM in the toll setting problem is of special interest when a welfare measure is involved. For example, in Wu et al. (2012), a pricing or credit scheme that maximizes the equity and the social welfare simultaneously in a general multimodal transportation network is described. Travelers' choices of modes and routes are represented by a nested logit model, and the resulting formulation is solved with an iterative derivative-free algorithm due to the presence of a numerical integration. Similarly, de Palma et al. (2018) propose a methodology to compute and compare (in terms of social welfare) optimal tolling systems in dollars and tokens or permits in the presence of static congestion when both demand (governed by a mixture of logit models) and capacity are stochastic. The benchmark is based on the optimum social welfare for each instrument obtained by solving a non-convex optimization problem.

Simulation-based optimization integrates optimization techniques into simulation analysis and revolves around methods that require the optimization of the net rewards (or costs) obtained from a random system (Gosavi et al., 2015). The problem at hand is usually solved to optimality (or near-optimality) with an iterative procedure that construct sequences of progressively better approximations to a *solution*, i.e., a point in the search-space that satisfies an optimality condition (Nguyen et al., 2014). This methodology has been lately considered in traffic assignment and toll pricing with a DCM (usually the logit model) representing the behavior of individuals. Gupta et al. (2020) propose an integrated framework that combines the optimization of network control strategies, which are solved with a genetic algorithm, with the generation of guidance information for real-time dynamic traffic assignment systems. The complex supply-demand relationship is characterized through a function named simulator that incorporates the behavioral response into the framework as a black box. Alternatively, Osorio and Atasoy (forthcoming) introduce an analytical network model formulated as a system of nonlinear equations (that can be efficiently evaluated with standard solvers) which is embedded within a metamodel simulation-based optimization algorithm. This enables the algorithm not to treat the simulator as a black box. The results show that the analytical structural information makes the proposed algorithm robust to the simulator's stochasticity.

Along similar lines, another research direction that has received some attention in the last years for the disaggregate representation of the demand is Markov chain choice models. These models formulate the substitution behavior of customers by state transitions in a Markov chain. In assortment optimization, Blanchet et al. (2016) address the problem of identifying the *right* model capturing substitution patterns with Markov chain choice models, since they provide a simultaneous approximation for all DCM based on random utility. The authors conclude that these provide a good approximation to the true choice probabilities under mild assumptions, and they feature a polynomial-time algorithm that compute the optimal assortment for a mixture of logit models. Likewise, Feldman and Topaloglu (2017) provide a LP formulation for the assortment optimization, single resource RM, and network RM problems under the Markov chain choice model. The computed experiments show that the logit model might be preferable because it prevents the fact that the Markov chain choice model may suffer from overfitting and it is able to compute the mean utility of a product without having to reestimate the parameters of the choice model. In traffic assignment, Ahipaşaoğlu et al. (2019) further develop on the Markovian traffic equilibrium model proposed by Baillon and Cominetti (2008) by relaxing the assumption that the distribution of the random utilities is known. They describe a distributionally robust Markovian choice model under the assumption that the joint distribution of the error terms is not known, and provide an equivalent convex optimization reformulation and an efficient solution algorithm to obtain the link choice probabilities and equilibrium flows.

The review of the literature shows that superior design and planning decisions can be obtained when the interplay between such decisions and the demand is explicitly represented. For this interplay to be captured, the operator's decisions need to be part of the set of explanatory variables of the DCM under consideration, i.e., they are included in the utility functions. In the case of facility location problems, if the only decisions are the location decisions, they might be exogenous to the utility functions (in contrast to other variables such as the price of a product or the frequency of a service). This allows to preprocess the choice probabilities for given values of the location variables, which notably simplifies the formulation and the associated solution methodologies. More generally, and due to the complexity of the resulting formulations, the proposed models include a restrained set of decision variables (e.g., only the price in RM or the toll in network pricing), even though additional supply-related decisions might as well be interconnected with the behavior of individuals and could therefore be simultaneously optimized.

Concerning the demand model, various authors have made some simplistic assumptions on the DCM in order to come up with tractable and more efficient solutions (Vulcano et al., 2010, Liu and Van Ryzin, 2008) although they might be inappropriate in reality. This is why the logit model is broadly considered regardless of its limitations. Indeed, more complex DCM lead to optimization problems that are more difficult to handle and that might not even have a closed form. The demand representation is usually integrated in the optimization model by embedding the corresponding choice probabilities. In general terms, an efficient solution of the problems is pursued by linearizing the associated formulation or by developing a tailored solution methodology, whose effectiveness might be limited to instances of small and moderate size. Other approaches that rely on simulation-based optimization have been recently proposed, both with DCM and alternative demand representations provided by Markov chain choice models.

In conclusion, there is an opening for a general methodology that bridges the gap between (complex) DCM and optimization models by allowing to accommodate a disaggregate demand representation that captures the interplay between the demand and the supply-related decisions. In this thesis, we investigate the integration of DCM in MILP with the end of enhancing the flexibility and tractability of the resulting formulations. The optimization problems that describe the operator's decisions aim at optimizing an aggregate performance of the system on such decisions while incorporating the corresponding demand responsiveness. The demand being provided by a DCM based on the random utility principle results in uncertainty within the optimization problem. We handle this uncertainty with the generation of possible outcomes (scenarios) of the random component associated with DCM. Hence, the general philosophy of the modeling framework here proposed is related to stochastic programming. For each scenario, and because of utility maximization as the decision rule, we assume that each individual solves an optimization problem in order to come up with a choice. The introduced framework is therefore also related to bilevel optimization since multiple problems at the lower level (utility maximization) are nested within another problem at the upper level (operator's decisions).

1.3 Research objectives and scientific contributions

As discussed in Section 1.2, the goal of this thesis is to develop a framework that embeds a disaggregate demand representation provided by a DCM in a MILP formulation that describes the supply-related decisions to be optimized. With this aim in mind, the main scientific objectives are the following:

1. **Model:** to propose operational optimization models that potentially accommodate any advanced DCM. This allows to provide a disaggregate representation of the demand that is able to capture the interactions between the individuals' behavior and the design and planning decisions to be optimized.
2. **Algorithm:** to design algorithms that exploit the structure of the model in order to address its complexity and enhance its tractability and potential scaling-up.
3. **Application:**
   (a) to give access to the huge literature on DCM that have been calibrated and validated by experts on real data and introduce them as such in the model, and
   (b) to apply the modeling framework in realistic contexts as a proof-of-concept.

This thesis contributes to these three objectives as follows. On the modeling side, we propose a choice-based optimization framework that allows to integrate DCM in MILP formulations. With respect to the algorithm, we rely on Lagrangian decomposition to exploit the decomposable structure of the model. Finally, we show the relevance of the proposed methodology in specific contexts where the use of advanced demand models is of special importance. The main scientific contributions of this thesis include:

1. **Model:**
   (a) We avoid the use of the probability expressions of DCM by specifying the associated preference structure directly in terms of the utility functions (structural equations of DCM).
   (b) We rely on simulation in order to tackle the stochastic nature of the utilities by drawing from the distribution of the associated random component, which enables to incorporate DCM in a MILP formulation.
   (c) We capture the interactions between the individuals' behavior and the supply-related decisions to be optimized by using a mixed-integer linear formulation of the equations and behavioral preference structure of DCM.
   (d) We express capacity constraints with respect to the supply-related decisions in terms of the variables of the choice-based optimization framework.
   (e) We define a flexible modeling framework that allows to accommodate different aggregation levels when it comes to group individuals with homogeneous behavior and to explicitly control the trade-off between model accuracy and tractability thanks to simulation.
   (f) We provide a linear approximation of the consumer surplus in terms of the expected maximum utility, which allows to derive measures of social welfare for public policy analysis.
2. **Algorithm:**
   (a) We provide an assessment of the viability of various decomposition strategies based on Lagrangian relaxation that are commonly used in practice with respect to the choice-based optimization framework, as the interrelations between some of the variables and constraints make such approaches unsuitable to this case.
   (b) We develop and test a heuristic approach based on Lagrangian decomposition in terms of a concrete MILP formulation (RM problem) that allows to rapidly generate upper bounds on the optimal value of the objective function, at the same time that feasible solutions that are close enough to the optimal solutions are obtained.
3. **Application:**
   (a) We show that the choice-based optimization framework can be directly applied with complex DCM borrowed from the literature without modifications, i.e., only the original explanatory variables of the DCM are considered.
   (b) We illustrate the usage and extent of the choice-based optimization framework with two applications: the maximization of the profit obtained by an operator that offers services to a market, and the maximization of a mesure of social

welfare by a public authority that decides on the pricing and design variables of a transportation system.

(c) We test the resulting formulations with different case studies that consider a variety of complex DCM, namely a nested logit model, a mixture of logit models and an integrated choice and latent variable (ICLV) model.

1.4 Structure of the thesis

This thesis is structured as follows.

Chapter 2 presents the choice-based optimization framework that allows to incorporate DCM in MILP formulations, and provides a comprehensive case study that shows its application and flexibility for the profit maximization problem.

This chapter borrows from a paper that is currently under review. Preliminary versions of the paper are published as

Pacheco, M., Sharif Azadeh, S., Bierlaire, M., and Gendron, B. (2017b). Integrating advanced discrete choice models in mixed integer linear optimization. Technical Report TRANSP-OR 170714, Transport and Mobility Laboratory, ENAC, EPFL.

Pacheco, M., Sharif Azadeh, S., and Bierlaire, M. (2016). A new mathematical representation of demand using choice-based optimization method. In *16th Swiss Transport Research Conference*.

Preliminary stages of this work have been presented in the following conferences:

- 15th International Conference on Travel Behavior Research, University of California Santa Barbara, July 17, 2018, Santa Barbara, CA, USA
- Seminario en el Departamento de Ingeniería de Transporte y Logística, Pontificia Universidad Católica de Chile, October 13, 2016, Santiago de Chile, Chile
- Seminario ISCI, Universidad de Chile, October 12, 2016, Santiago de Chile, Chile
- 5th Symposium of the European Association for Research in Transportation (hEART), Delft University of Technology, September 15, 2016, Delft, Netherlands
- Ninth Triennial Symposium on Transportation Analysis (TRISTAN IX), June 14, 2016, Oranjestad, Aruba
- 16th Swiss Transport Research Conference (STRC), May 18, 2016, Monte Verità, Ascona, Switzerland

Chapter 3 proposes an algorithmic solution approach that relies on Lagrangian decomposition in order to generate near-optimal solutions in a much more reduced computational time. The procedure is illustrated with the RM problem.

Preliminary ideas related to this chapter are published as

> Pacheco, M., Lurkin, V., Gendron, B., Sharif Azadeh, S., and Bierlaire, M. (2018). Lagrangian relaxation for the demand-based benefit maximization problem. In *18th Swiss Transport Research Conference (STRC), Ascona, Switzerland.*
>
> Pacheco, M., Sharif Azadeh, S., Bierlaire, M., and Gendron, B. (2017a). Integrating advanced demand models within the framework of mixed integer linear problems: a lagrangian relaxation method for the uncapacitated case. In *17th Swiss transport research conference (STRC), Ascona, Switzerland.*

Preliminary stages of this work have been presented in the following conferences:

- Annual international conference of the German Operations Research Society (OR 2018), Université libre de Bruxelles, September 14, 2018, Brussels, Belgium
- 7th Symposium of the European Association for Research in Transportation (hEART), National Technical University of Athens, September 06, 2018, Athens, Greece
- Workshop on Discrete Choice Models 2018, EPFL, June 22, 2018, Lausanne, Switzerland
- 16th Swiss Operations Research Days, Universität Bern, June 12, 2018, Bern, Switzerland
- 8th Swiss Transport Research Conference (STRC), May 17, 2018, Monte Verità, Ascona, Switzerland
- Annual conference of the Belgian Operational Research Society (ORBEL 32), HEC Liège, February 02, 2018, Liège, Belgium
- Annual international conference of the German Operations Research Society (OR 2017), Freie Universität Berlin, September 07, 2017, Berlin, Germany
- 21st Conference of the International Federation of Operational Research Societies, IFORS, July 19, 2017, Québec City, Canada
- 15th Swiss Operations Research Days, Université de Fribourg, June 30, 2017, Fribourg, Switzerland
- Workshop on Discrete Choice Models 2017, EPFL, June 22, 2017, Lausanne, Switzerland

Chapter 4 describes an application of the choice-based optimization framework in the context of pricing and design of a transportation system when a measure of social welfare is to be maximized. The methodology is illustrated with two case studies that rely on different DCM.

Preliminary ideas related to this chapter are published as

Pacheco, M., Sharif Azadeh, S., and Bierlaire, M. (2019). Passenger satisfaction maximization within a demand-based optimization framework. In *19th Swiss Transport Research Conference (STRC), Ascona, Switzerland.*

Preliminary stages of this work have been presented in the following conference:

- 19th Swiss Transport Research Conference (STRC), May 16, 2019, Monte Verità, Ascona, Switzerland

Chapter 5 summarizes the main findings and contributions of this thesis and discusses some avenues for future research.

2

Choice-based optimization framework

This chapter is based on the technical report (currently under review in the journal `Transportation Research Part B: Methodological`)

`Pacheco, M., Sharif Azadeh, S., Bierlaire, M., and Gendron, B. (2017b). Integrating advanced discrete choice models in mixed integer linear optimization. Technical Report TRANSP-OR 170714, Transport and Mobility Laboratory, ENAC, EPFL.`

The work has been performed by the candidate under the supervision of Prof. Shadi Sharif Azadeh and Prof. Michel Bierlaire and the collaboration of Prof. Bernard Gendron.

2.1 Introduction

Discrete choice models (DCM) are the state-of-the-art of demand modeling at the disaggregate level. Among various advantages, these models enable to capture the heterogeneity of tastes and preferences in high detail, and combined with the more and more available individual-based datasets, they allow to predict a wide range of behaviors in a great deal of contexts. Unfortunately, their integration in optimization models formulating supply-related decisions that require a disaggregate demand representation is a challenging task, often making DCM not to be considered or, when considered, to present the supply-related decisions as exogenous, and therefore not capturing the interactions.

The main reason for this lack of integration is the different types of focus on both sides. On the one hand, optimization models focus on tractability and availability of solution techniques. This is why mixed-integer linear problems (MILP) represent a significant share of the models reported in the literature. On the other hand, DCM focus on behavioral realism, which results in complex mathematical formulations, certainly not linear or even convex, that are difficult to embed in MILP.

The objective of this chapter is to define a methodology that allows to incorporate a disaggregate demand representation that interacts with the decisions to be optimized. To this end, we propose a general framework that allows to integrate a DCM based on the random utility principle in MILP. The only condition that we impose on the DCM is that the decision variables of the MILP that have an impact on the behavior of individuals, and therefore are also present in the DCM (such as the price of a service or the frequency of a transportation mode), appear linearly in the utility function.

The key idea is to express the demand in terms of the utility function (instead of the probability functions of DCM), and to rely on simulation to overcome the stochastic nature of the associated random component. Notwithstanding the potentially large size of the formulation, the trade-off between model accuracy and tractability can be explicitly controlled by the modeler thanks to simulation.

The contribution of this research is twofold. First, we determine a mixed-integer linear formulation of a DCM that can be embedded in MILP. For the sake of illustration, we define the problem of an operator that wants to maximize its profit, but other MILP models with different objectives and/or restrictions could be specified (see Chapter 4 for a detailed application on welfare maximization). Second, we show that the framework can be directly applied with an existing DCM from the literature without modifications, i.e., only the original variables that appear in the DCM are considered. We also show the flexibility of the framework by adapting the resulting formulation to test different contexts (e.g., price differentiation by population segmentation) and assumptions (e.g., grouping of individuals with homogeneous behavior).

The remainder of the chapter is organized as follows. Section 2.2 describes the modeling framework and Section 2.3 depicts the profit maximization problem. The case study used as a proof-of-concept is detailed in Section 2.4. Finally, some concluding remarks and future research directions are discussed in Section 2.5.

2.2 Modeling framework

We assume that the demand is characterized by a DCM, whose parameters are estimated at a preprocessing stage, and that the decisions to be made by an operator are governed by an optimization model, more precisely a MILP. We consider three types of variables within the framework: the exogenous variables explaining the choice and not involved in the optimization model $x^d \in \mathbb{R}^D$, the exogenous optimization variables not involved in the choice model $x^s \in \mathbb{R}^S$, and the endogenous variables $x^e \in \mathbb{R}^E$, which are involved in both the choice and the optimization models, i.e., they are operator's decisions that also appear in the utility functions of the behavioral model. Depending on the specifications of the model, these variables can be restricted to take integer or binary values. The exogenous variables appear in one of the two models, but not in both. The endogenous variables are present in both, and characterize the interactions. For the definition of the optimization model, they are assumed to be bounded:

$$\ell^e \leq x^e \leq m^e, \tag{2.1}$$

where $\ell^e \in \mathbb{R}^E$ is the vector of lower bounds on x^e and $m^e \in \mathbb{R}^E$ is the vector of upper bounds.

A typical example of an endogenous variable is the price of a service. The operator decides on a price in order to maximize its revenue, and the individual reacts to the price to decide if they buy the service or not. If the price is too high, few individuals will access the service, and a low revenue will be generated. If it is too low, many individuals will use the service, but the generated revenue will also be low. This example is treated extensively in Section 2.4. Other examples of endogenous variables are the schedule of an event (e.g., departure of a train) and the capacity of a facility (e.g., number of coaches in a train).

2.2.1 The discrete choice model

The choices of individuals are modeled with a DCM. The set of all potential alternatives is called the choice set and is denoted by $\mathcal{C}$. The alternatives in $\mathcal{C}$ are indexed by i. For each alternative i, we denote by $c_i \geq 1$ its capacity, that is, the maximum number of individuals who can choose it. We allow for a population of N individuals, indexed by $n \geq 1$. Generally, it is impossible to have access to the full population, and a sample must be used. A synthetic population, which is constructed by combining different data sources, is also convenient here (Farooq et al., 2013). The following description, based on the full population, can be easily adapted to a representative sample.

The choice set of two different individuals may not be the same. The choice set of individual n is denoted by $\mathcal{C}_n \subseteq \mathcal{C}$. It contains the alternatives considered by and offered to individual n, as in some cases some alternatives may not be offered to some individuals for certain reasons. For instance, from a profit maximization point of view, a service that is not profitable will not be proposed. These decisions are modeled with the binary variables y_{in}, which are 1 if alternative i is considered by and offered to individual n, and 0 otherwise (see Section 2.2.2). These variables are endogenous, i.e., they belong to the vector x^e. Therefore, the set of offered alternatives is flexible in the sense that it is possible not to propose some alternatives to some specific individuals, or in a more practical manner, groups of individuals. This feature allows the decision-maker to investigate different marketing solutions and business models.

The preference structure of individuals is represented with a utility function, which associates a score with each alternative $i \in \mathcal{C}_n$. This utility is defined as

$$U_{in}(x^d_{in}, x^e_{in}; \varepsilon_{in}) = V_{in}(x^d_{in}, x^e_{in}) + \varepsilon_{in}, \tag{2.2}$$

where $V_{in} : \mathbb{R}^{D+E} \to \mathbb{R}$ is the systematic part of the utility function, that includes everything that can be modeled by the analyst, and ε_{in} is the random component, that captures everything that has not been included explicitly in the model and is independent of the exogenous demand variables x^d_{in} and endogenous variables x^e_{in} associated with alternative i and individual n. As ε_{in} is a random variable, $U_{in}(x^d_{in}, x^e_{in}; \varepsilon_{in})$ is also a random variable.

The behavioral assumption is that individual n chooses alternative i if the corresponding utility is the largest within the choice set $\mathcal{C}_n$ (Manski, 1977). We assume that each individual chooses one and only one alternative. The probability that individual n chooses alternative i within the choice set $\mathcal{C}_n$ is

$$P_n(i|x^d_{in}, x^e_{in}) = \Pr(U_{in}(x^d_{in}, x^e_{in}; \varepsilon_{in}) \geq U_{jn}(x^d_{jn}, x^e_{jn}; \varepsilon_{jn}), \forall j \in \mathcal{C}_n). \tag{2.3}$$

Throughout the thesis, it is assumed that V_{in} is linear in the endogenous variables x^e_{in}. This is not required as such for the derivation of the choice model, but important in our context for its integration in MILP. For this reason, the deterministic term in (2.2) is written as

$$V_{in}(x^d_{in}, x^e_{in}) = \sum_k \beta_{ink} x^e_{ink} + g^d_{in}(x^d_{in}), \tag{2.4}$$

where x^e_{ink} is the k-th endogenous variable associated with alternative i and individual n and β_{ink} are the associated coefficients. The functions $g^d_{in} : \mathbb{R}^D \to \mathbb{R}$ do not need to be linear since the variables x^d_{in} are not involved in the optimization model, i.e., $g^d_{in}(x^d_{in})$ is a quantity that can be preprocessed. The parameters β_{ink} of the DCM are previously estimated (outside of the optimization scheme), i.e., they are not variables of the model. Nevertheless, the framework here described can be adapted to address the parameter estimation problem. We refer the reader to Fernández-Antolín et al. (2017) for further details.

Operational choice models are obtained by assuming a distribution for the error term ε_{in}. For example, the logit model is obtained by assuming that they are independent and identically distributed (across both i and n), with an extreme value distribution. It can be shown that, for the logit model, (2.3) is written as

$$P_n(i|x_{in}^d, x_{in}^e) = \frac{y_{in} e^{V_{in}(x_{in}^d, x_{in}^e)}}{\sum_{j \in \mathcal{C}_n} y_{jn} e^{V_{jn}(x_{jn}^d, x_{jn}^e)}}. \tag{2.5}$$

Advanced DCM, which aim at relaxing the unrealistic assumptions associated with the logit model and have shown a better prediction power, can also be accommodated within this framework. In the case study described in Section 2.4, a mixed logit model is considered. These models can be derived under a variety of behavioral specifications whose choice probabilities take a specific functional form, and each derivation provides a particular interpretation. For some derivations of the mixed logit model (and specifically the one in Section 2.4), the deterministic part of the utility specification of the standard logit model is generalized by allowing one or some of the coefficients β_{ink} in (2.4) to be randomly distributed across the population, which captures heterogeneity among individuals. The vector of coefficients β_{nk} of individual n is therefore a random vector, with probability density function $f(\beta_k|\theta)$, θ being the parameters of the distribution of β_{nk}, such as their mean and variance. The probability that individual n chooses alternative i is given by the standard logit formula conditional on β_{nk}. As β_{nk} is distributed, the (unconditional) choice probability (2.3) is the integral of the logit formula over the density of β_{nk}:

$$P_n(i|x_{in}^d, x_{in}^e) = \int \frac{y_{in} e^{V_{in}(x_{in}^d, x_{in}^e; \beta_{nk})}}{\sum_{j \in \mathcal{C}_n} y_{jn} e^{V_{jn}(x_{jn}^d, x_{jn}^e; \beta_{nk})}} f(\beta_k|\theta) d\beta_k. \tag{2.6}$$

Latent factors, such as personal attitudes and perceptions, also allow for a more realistic representation of the behavior inherent in the choice process. They have been integrated in DCM through two main approaches: models with latent variables, which explicitly model the unobserved psychological characteristics of the individual (see Section 4.4 for an application), and latent class models, which assume that the population can be probabilistically segmented into discrete groups that have different choice behaviors. A special case of the latter corresponds to the DCM with latent choice sets (Ben-Akiva and Boccara, 1995), which model individual choice behavior as a two-stage process consisting of choice set generation first followed by a choice from the resulting given choice set. This enables to incorporate an explicit probabilistic representation of the availability of the alternatives, instead of assuming them as given, as is the case in standard DCM.

The expected demand for each alternative $i \in \mathcal{C}$ is then given by

$$D_i = \sum_{n=1}^{N} P_n(i|x_{in}^d, x_{in}^e). \tag{2.7}$$

We note that (2.7) can be used to derive, for example, the expected gain obtained from alternative i: $G_i = p_i D_i$, where p_i denotes the price associated with alternative i (see Section 2.3). Since (2.5) and (2.6) are non-linear as a function of the variables x^e_{in}, so is the associated expression of the demand (2.7). Moreover, price is typically an endogenous variable, which makes the formulation of the expected gain even more complex.

We illustrate the non-linearity and non-convexity of the previous expressions by means of an example for a simple logit model. Consider a choice set composed of two alternatives: the same service offered by a certain operator (i) and its competitor (j). A population of $N = 100$ individuals consists of two groups with different behavior: group 1 (2/3 of the population) and group 2 (1/3 of the population). The systematic term of the utility function for alternative i is defined as $V_{ig} = \beta_g p_i + \mathrm{s}_{ig}$, where g denotes the group, β_g the price sensitivity of an individual in group g, p_i the price to access service i and s_{ig} is the term associated with service i and group g capturing the socioeconomic characteristics of interest. We assume that group 1 is highly sensitive to price ($\beta_1 = -10$) and has an intrinsic preference towards alternative i ($\mathrm{s}_{i1} = 3$), whereas group 2 has a lower price sensitivity ($\beta_2 = -1$) and does not have a preference for alternative i ($\mathrm{s}_{i2} = 0$). Furthermore, we assume that no information is available about the competitor, and therefore we define $V_{jg} = 0$ for $g = 1, 2$.

Figure 2.1 shows the expected revenue for service i obtained from the two groups (separately), and the sum of the two, as a function of price. The revenue function is unimodal within each group, while the total gain curve is bi-modal: the first local optima is reached when both groups are attracted to service i because of the low price, and the second one is related to group 2 only, as group 1 has decided to leave the market due to the high price.

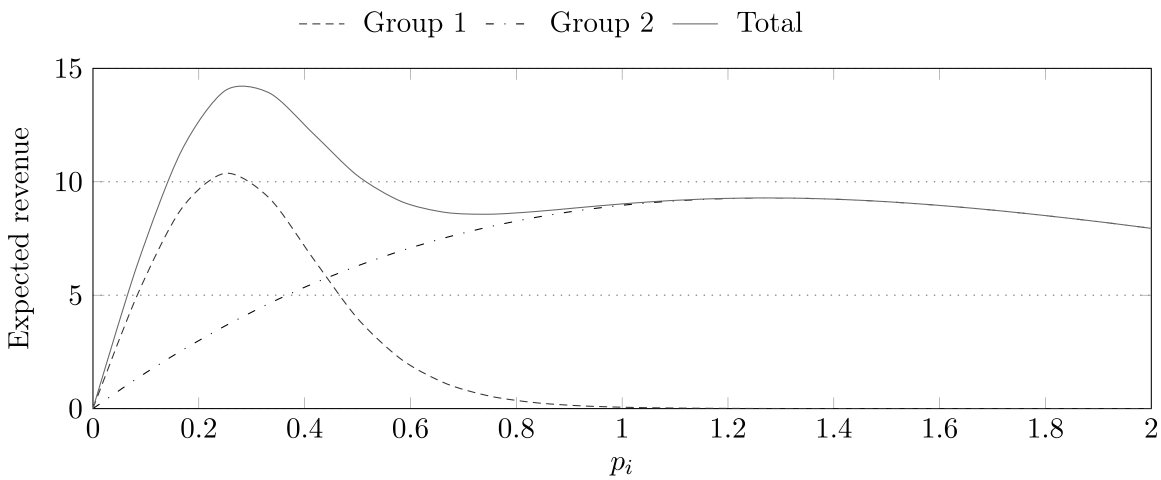

Figure 2.1: Expected revenue obtained from group 1, group 2 and total as a function of the price of alternative i (p_i)

In general, when real models involving heterogeneity in the population are considered, the associated objective functions are multi-modal, as we have illustrated in Figure 2.1 for a simple logit model with only two groups in the population. In this thesis, we define a framework that avoids the use of the probability expressions by specifying the DCM directly in terms of the utility functions.

Concerning the interactions between the demand and the supply-related decisions, we note that the binary variables y_{in} are endogenous to the formulation, as they belong to both the DCM (see probability expressions (2.5) and (2.6)) and the optimization model. However, they are exogenous to the utility functions (they do not appear in $V_{in}(x^d_{in}, x^e_{in})$). The presence of endogenous variables in the utility function makes the formulation more complex, since $V_{in}(x^d_{in}, x^e_{in})$ cannot be preprocessed. We propose a general approach that allows for both types of endogenous variables, as described next in Section 2.2.2.

2.2.2 Simulation-based linearization

Commonly, when a probabilistic model like the one introduced in Section 2.2.1 wants to be integrated in MILP, it is tackled either with simulation (to approximate the probability expressions, especially when they do not have a closed form) or with linearization of the non-linear probability expressions. We rely on simulation to address the stochasticity of the random component in the utility function (2.2). This enables to approximate the expected demand in terms of the utility function with a mixed-integer linear formulation based on the sample average approximation (SAA) principle in the space of the utilities.

For each ε_{in} in (2.2), we generate R simulation draws $\xi_{in1},\ldots,\xi_{inR}$ from its distribution (e.g., Gumbel, normal) outside of the optimization procedure (known as exterior approach in stochastic programming), where ξ_{inr} denotes the r-th draw. Each draw can be seen as an independent behavioral scenario. We notice that variance reduction techniques (e.g., linear control random variables method, importance sampling) could be used to enhance convergence of the SAA estimators.

Once the draws have been generated, for each scenario r, we obtain the utility associated with alternative i by individual n, which is denoted by $U_{inr}(x^d_{in}, x^e_{in})$, or simply U_{inr}. For the specification (2.4), we have

$$U_{inr} = V_{in} + \xi_{inr} = \sum_k \beta_{ink} x^e_{ink} + f^d_{in}(x^d_{in}) + \xi_{inr}, \qquad \forall i \in \mathcal{C}_n, n, r. \tag{2.8}$$

As the variables x^e_{ink} are bounded (see (2.1)), and the values $f^d_{in}(x^d_{in})$ are given, lower and upper bounds on U_{inr} can be derived. They are denoted by ℓ_{inr} and m_{inr}:

$$\ell_{inr} \leq U_{inr} \leq m_{inr}, \qquad \forall i \in \mathcal{C}_n, n, r. \tag{2.9}$$

Notice that the idea of relying on scenarios to represent uncertainty is exploited in stochastic programming, and in particular in one of the basic modeling approaches known as the multi-stage stochastic programming model (Ruszczyński and Shapiro, 2003). This general structure is usually implemented for two stages, known as the two-stage model, and is widely used in stochastic programming. In a standard two-stage model, the decision variables are divided into two groups: first-stage and second-stage variables. First-stage variables are decided upon before the actual realization of the random parameters such that the expected value of an objective function (which in turn is the optimal value of the second-stage optimization problem) is optimized. Once the uncertain events have unfolded, further design or operational adjustments can be made through values of the second-stage (or alternatively called recourse) variables at a particular cost.

To numerically solve the two-stage problem, the vector of random parameters is often assumed to have a discrete distribution with a finite number of possible outcomes (scenarios) with respective probability masses, which yields a deterministic equivalent formulation of the problem. The number of scenarios should be relatively modest such that the deterministic equivalent formulation can be solved within reasonable computational effort. If the total number of scenarios is very large (or even infinite), the scenario set can be reduced to a manageable size by generating a sample of replications (draws) of the random vector. The expectation function of the formulation is then approximated with the SAA method, which allows to solve the problem using deterministic algorithms.

In this case, by generating R draws (scenarios) of the random component $\varepsilon_{in}, \forall i \in \mathcal{C}_n, n$, we associate a deterministic utility function with each scenario (see (2.8)). These utilities are considered to determine a choice for each individual and scenario, which allows to approximate the individual choice probabilities, and consequently the expected demand, with the SAA method. A detailed description is provided next.

Availability of alternatives

An alternative may be unavailable for three reasons. First, the operator decides that the alternative is not made available to individual n. This decision is modeled with the binary variables y_{in} introduced in Section 2.2.1, which are formally defined as

$$y_{in} = \begin{cases} 1 & \text{if alternative i is offered to individual n,} \\ 0 & \text{otherwise,} \end{cases} \qquad \forall i \in \mathcal{C}, n. \tag{2.10}$$

Second, an alternative might not be considered by the individual ($i \notin \mathcal{C}_n$, where $\mathcal{C}_n$ comes from external data and represents the set of alternatives considered by individual n). This exogenous decision can be explicitly included in the MILP by adding the following constraint:

$$y_{in} = 0, \qquad \forall i \notin \mathcal{C}_n, n. \tag{2.11}$$

Third, the alternative may be unavailable because its capacity has been reached. This type of unavailability is more complex to model, as it is not a direct decision as such, but the result of the decisions of other individuals. Note that in this framework this can vary from one draw to the next. Indeed, an alternative might be more attractive in one scenario, generating more demand than its capacity, and less attractive in another one.

We model the availability of alternative i to individual n in scenario r with the binary variables y_{inr}. Note that the variables y_{in} and y_{inr} are related as follows:

$$y_{inr} \leq y_{in}, \qquad \forall i \in \mathcal{C}, n, r, \tag{2.12}$$

which implies that alternative i is not available at scenario level ($y_{inr} = 0$) if $y_{in} = 0$.

Briefly, the variables y_{in} and y_{inr} model three different situations. If the alternative is not offered to or considered by individual n ($y_{in} = 0$), then $y_{inr} = 0\ \forall r$ due to constraints (2.12). If the alternative is offered and considered ($y_{in} = 1$), and there is still room for individual n, then $y_{inr} = 1$. However, if the capacity of the alternative has been reached, $y_{inr} = 0$ (even if $y_{in} = 1$).

Discounted utility

The behavioral assumption states that the individual selects the alternative associated with the largest utility. To avoid that an unavailable alternative is related to the highest utility, we introduce the concept of discounted utility, which is the utility itself when the alternative is available, and a low value otherwise. The discounted utility associated with alternative i by individual n in scenario r is defined as

$$z_{inr} = \begin{cases} U_{inr} & \text{if } y_{inr} = 1, \\ \ell_{nr} & \text{if } y_{inr} = 0, \end{cases} \qquad \forall i \in \mathcal{C}, n, r, \tag{2.13}$$

where $\ell_{nr} = \min_{j \in \mathcal{C}_n} \ell_{jnr}$ is the smallest lower bound across all alternatives.

The linear formulation of (2.13) is given by

$$\begin{aligned} \ell_{nr} &\leq z_{inr}, && \forall i \in \mathcal{C}, n, r, && (2.14)\\ z_{inr} &\leq \ell_{nr} + M_{inr} y_{inr}, && \forall i \in \mathcal{C}, n, r, && (2.15)\\ U_{inr} - M_{inr}(1 - y_{inr}) &\leq z_{inr}, && \forall i \in \mathcal{C}, n, r, && (2.16)\\ z_{inr} &\leq U_{inr}, && \forall i \in \mathcal{C}, n, r, && (2.17) \end{aligned}$$

where

$$M_{inr} = m_{inr} - \ell_{nr}. \tag{2.18}$$

To verify that (2.13) is equivalent to constraints (2.14)–(2.17) we consider two cases. If $y_{inr} = 0$, constraints (2.14)–(2.15) impose that $z_{inr} = \ell_{nr}$ and constraints (2.16)–(2.17) are always satisfied (using (2.18) and the definition of ℓ_{nr}, respectively). If $y_{inr} = 1$, constraints (2.16)–(2.17) impose that $z_{inr} = U_{inr}$, and constraints (2.14)–(2.15) are always satisfied (using the definition of ℓ_{nr} and (2.18), respectively).

Choice

The choice of individual n in scenario r is characterized by the following binary variables:

$$w_{inr} = \begin{cases} 1 & \text{if } i \text{ is chosen,} \\ 0 & \text{otherwise,} \end{cases} \qquad \forall i \in \mathcal{C}, n, r. \tag{2.19}$$

As each individual is choosing exactly one alternative, we impose

$$\sum_{i \in \mathcal{C}} w_{inr} = 1, \qquad \forall n, r. \tag{2.20}$$

Moreover, since an alternative that is not available cannot be selected, we add the following constraint:

$$w_{inr} \leq y_{inr}, \qquad \forall i \in \mathcal{C}, n, r. \tag{2.21}$$

In terms of the discounted utility, the chosen alternative corresponds to the one with the highest discounted utility, which is represented by the continuous variables U_{nr} defined as

$$U_{nr} = \max_{i \in \mathcal{C}} z_{inr}, \qquad \forall n, r. \tag{2.22}$$

The linear formulation of (2.22) is given by

$$z_{inr} \leq U_{nr}, \qquad \forall i \in \mathcal{C}, n, r, \tag{2.23}$$

$$U_{nr} \leq z_{inr} + M_{nr}(1 - w_{inr}), \qquad \forall i \in \mathcal{C}, n, r, \tag{2.24}$$

where

$$M_{nr} = m_{nr} - \ell_{nr} \tag{2.25}$$

is the difference between the largest upper bound and the smallest lower bound, where the largest upper bound is defined as $m_{nr} = \max_{j \in \mathcal{C}_n} m_{jnr}$.

To prove the equivalence between definition (2.22) and the formulation (2.23)–(2.24), we consider two cases. If $w_{inr} = 0$, constraints (2.23) are consistent with (2.22), and constraints (2.24) are always verified (using (2.25)). If $w_{inr} = 1$, constraints (2.23)–(2.24) impose that $U_{nr} = z_{inr}$, which means that alternative i is associated with the highest discounted utility. Notice that it may happen that two alternatives correspond to the highest utility. In this case, constraints (2.20) guarantee that only one of them is chosen, and the actual choice will be governed by the specific optimization problem, which is not behaviorally realistic. However, thanks to simulation, such an issue is happening sufficiently rarely to be ignored.

Note that constraints (2.17) and (2.23) are equivalent. We have decided to keep both in the model as it has proven to be computationally more efficient (by performing some preliminary tests that ignore one of the two constraints at a time). Furthermore, the characterization of the so-called big M constraints (constraints (2.14)–(2.17) and (2.23)–(2.24)) is tailored to our problem. Indeed, the values of the M constant defined in (2.18) and (2.25) are the tightest possible values for the associated constraints.

Expected demand

The complexity of the probability distributions of the random variables involved in the DCM and their correlation structure are irrelevant in this context as long as it is possible to draw from these distributions (which is performed at a preprocessing stage). Given an independent and identically distributed sample $\xi_{in1},\ldots,\xi_{inR}$ of the random variable ε_{in}, the choice variables w_{inr} allow to count the number of times that the behavioral assumption associated with $P_n(i|x^d_{in}, x^e_{in})$ (see (2.3)) in terms of U_{inr} is met, which enables to compute the sample average

$$\frac{1}{R}\sum_{r=1}^{R} w_{inr}. \tag{2.26}$$

Hence, as a consequence of the law of large numbers, the relative frequency calculated in (2.26) provides an estimation of $P_n(i|x^d_{in}, x^e_{in})$, and the total expected demand of alternative $i \in \mathcal{C}$ is approximated by aggregating such estimates across individuals:

$$D_i \approx \frac{1}{R}\sum_{r=1}^{R}\sum_{n=1}^{N} w_{inr}. \tag{2.27}$$

Notice that the law of large numbers is valid if the expectation of the probability distribution associated with the error component ε_{in} is finite, which is the case of the operational DCM employed in practice.

Capacity allocation

If the demand for alternative i is larger than its capacity, it is necessary to decide which individuals have access to the alternative. We have decided to model it exogenously, using an externally defined priority list of individuals, similar to Binder et al. (2017). This list determines the way in which the individuals are processed (which does not necessarily coincide with the order in which individuals arrive), and as soon as capacity is reached, the remaining (unprocessed) individuals will not have access to the alternative. An individual has access to an alternative if all individuals before them in the list for which the alternative is offered have also access to it. Thus, the numbering of individuals reflects the priority list. Note that the construction of this priority list can consider various aspects of the relationship between the operator and the individuals, such as fidelity programs, VIP individuals, etc. It can also be randomly generated.

The capacity restrictions are expressed by the following constraints:

$$\sum_{m=1}^{n-1} w_{imr} \leq (c_i - 1)y_{inr} + (n-1)(1 - y_{inr}), \qquad \forall i \in \mathcal{C}, n > c_i, r, \tag{2.28}$$

$$c_i(y_{in} - y_{inr}) \leq \sum_{m=1}^{n-1} w_{imr}, \qquad \forall i \in \mathcal{C}, n > 1, r. \tag{2.29}$$

Constraints (2.28) forbid the access of individuals to a certain alternative when its capacity has been reached, whereas constraints (2.29) ensure the availability of the alternative when the capacity has not been exceeded. They can be verified by considering two cases when the alternative is offered to and considered by individual n (i.e., $y_{in} = 1$):

1. $\sum_{m=1}^{n-1} w_{imr} < c_i$, and
2. $\sum_{m=1}^{n-1} w_{imr} \geq c_i$.

In the first case, constraints (2.28) are always satisfied (for both $y_{inr} = 0$ and $y_{inr} = 1$), and constraints (2.29) force y_{inr} to be equal to 1 in order to be verified, which means that the number of individuals up to and including n who have chosen alternative i does not exceed c_i, so there is still room for individuals n. In the second case, constraints (2.29) are always satisfied, and constraints (2.28) imply $y_{inr} = 0$, which means that the capacity has been reached due to the choices of the individuals up to and including $n-1$, and even if the alternative is proposed to individual n by the operator, there is no room left for them.

Note that if alternative i is not offered to individual n by the operator or not considered by this individual ($y_{in} = 0$), the variables y_{inr} and w_{inr} are equal to 0 (due to constraints (2.12) and (2.21), respectively), and therefore constraints (2.28)–(2.29) are always satisfied.

2.2.3 Capacities as decision variables

In Section 2.2.2, we have assumed that the capacities c_i are given. However, the model can be easily extended to include capacity as a decision variable. In order to avoid the non-linearity that would appear in constraints (2.28)–(2.29) (due to the product of the capacity with the availability variables), we formulate capacity as follows.

For each decision variable c_i, a predefined list of Q feasible values for the capacity is proposed: $c_{i1}, \dots, c_{iQ}$. Then, alternative i is duplicated Q times, each instance being associated with the same utility function, but with a different capacity level. We define the binary variables y_{iq}, which take value 1 if alternative i is offered with capacity c_{iq}, and 0 otherwise. We also define the binary variables y_{iqn}, which represent the extension of the variables y_{in}, and therefore are equal to 1 if alternative i with capacity c_{iq} is offered to and considered by customer n, and 0 otherwise. It is sufficient to include the following constraints in the formulation:

$$\sum_{q=1}^{Q} y_{iq} \leq 1, \qquad \forall i \in \mathcal{C}, \tag{2.30}$$

$$y_{iqn} \leq y_{iq}, \qquad \forall i \in \mathcal{C}, q, n. \tag{2.31}$$

Constraints (2.30) guarantee that at most one of the duplicates is actually available. Note that it is still possible for the operator to decide not to offer alternative i at all. In that case, the sum on the left-hand side of (2.30) is equal to zero. Constraints (2.31) ensure that the variables y_{iqn} are set to 0 if alternative i is not offered with capacity c_{iq}.

The remaining variables described in Section 2.2.2 need to be extended to account for the capacity levels. We can simplify the introduced notation by redefining $\mathcal{C}$ as the set of the duplicates of the original alternatives. In this case, the dimension of $\mathcal{C}$ would be JQ, where J is the number of original alternatives.

2.2.4 Size of the mixed-integer linear formulation

The number of constraints comprised in this specification (when capacity is a decision variable) is of the order of $JNRQ$. In real applications, where the number of individuals can be large, this comes with a high computational price. To reduce the size of the model, individuals can be grouped into classes of homogeneous behavior (see Section 2.4.2 and 2.4.3 for a concrete example), even though this modeling technique requires additional assumptions on how to handle the access of groups to the services in order to fulfill the capacity restrictions. Moreover, this formulation allows to explicitly model the trade-off between the number of draws and the accuracy of the approximation (see Section 2.4.1).

As pointed out in Section 2.2.2, the complexity of the probability distributions of the random variables involved in the DCM and their correlation structure is only affected by the number of draws, and not by the nature of the underlying distributions. This is a strength of the framework, that it is relevant for any existing complex DCM, and for other models to be developed in the future.

Finally, we highlight the fact that the tremendous development of advanced mathematical formulations and efficient algorithms allow to solve MILP models of gigantic sizes. Furthermore, the structure of the described formulation is particularly well suited for decomposition methods. This is extensively discussed in Chapter 3.

2.2.5 The optimization model

The generic mixed-integer linear formulation of a DCM presented in the previous sections can be integrated into an optimization model consisting of:

- a linear objective function $f^s : \mathbb{R}^{S+E+J} \to \mathbb{R}$ that relates the decision variables and the expected demand $D_R \in \mathbb{R}^J$ (for a given R) to an aggregate performance of the system:

$$f^s(x^s, x^e, D_R), \tag{2.32}$$

- a set of linear constraints that identifies the feasible configurations of the variables:

$$\begin{aligned} g_1(x^s, x^e, D_R) &= 0, && (2.33)\\ g_2(x^s, x^e) &= 0, && (2.34)\\ x^e_z &\in \mathbb{Z}^{E_z}, && (2.35)\\ x^s_z &\in \mathbb{Z}^{S_z}, && (2.36)\\ \ell^e \le x^e &\le m^e, && (2.1)\\ \ell^s \le x^s &\le m^s, && (2.37) \end{aligned}$$

  where $g_1 : \mathbb{R}^{S+E+J} \to \mathbb{R}^{I_1}$ and $g_2 : \mathbb{R}^{S+E} \to \mathbb{R}^{I_2}$ (with $I_1, I_2 \ge 0$ denoting the number of constraints) represent the constraints that do and do not involve the demand model, respectively, x^e_z and x^s_z correspond to the subsets of integer variables in x^e and x^s, respectively, and ℓ^s and m^s denote the lower and upper bounds on x^s, respectively.

We can prove the convergence (when R tends to infinity) of the sequence of optimal solutions obtained with the optimization problem defined by (2.1), (2.32)–(2.37) to an optimal solution of the same problem relying on the probability-based demand representation of DCM (as defined in (2.7)). In Appendix B we provide additional details on this convergence property and the corresponding proof.

This formulation can be employed to model numerous applications. In Section 2.3, we illustrate the usage of the choice-based optimization framework by considering a concrete application. More precisely, we define a profit maximization problem that allows us to characterize a specific objective function and feasible configuration of the variables.

2.3 Profit maximization problem

We consider a profit maximization problem to illustrate how the framework described in Section 2.2 can be used. This application is particularly interesting because the calculation of revenue involves a non-linearity that needs to be addressed, and can be found in many different contexts (e.g., airline RM, road tolling).

The operator aims at finding the best strategy in terms of pricing and capacity allocation to maximize its profit by selling services, each of them at a certain price and with a certain capacity, both to be decided. Regarding the cost of each service, we assume that it is composed of a fixed cost associated with operating the service and a variable cost associated with each unit of the service sold.

The market is composed of N customers, which are assumed to be heterogeneous and price elastic, in the sense that each costumer may have a different behavior and sensitivity towards price. The operator is considering a set $\mathcal{C}$ composed of J services, each of them representing a service potentially offered by the operator and its associated capacity level, as described in Section 2.2.3.

In a profit maximization context, we need to model competition. If we do not account for competitive services, customers are captive, and the problem becomes unbounded. Competitive services can be explicitly modeled in the choice set, or grouped into an opt-out option that captures customers leaving the market, either because they choose a competitor's service or because they do not choose anything at all. To keep the illustrative example simple, we consider the second approach. The main assumption is that the decisions of the competitors are given, and not adjusted as a consequence of the decisions of the operator. The opt-out option is denoted by $i = 0$, and it is always available to all customers, i.e., $0 \in \mathcal{C}_n \; \forall n$ (i.e., $y_{0n} = 1$).

We consider the price as the only endogenous variable (x^e) in the utility function (2.8). We define $p_{in} \in \mathbb{R}$ as the price that customer n must pay to access service $i \in \mathcal{C}_n \setminus \{0\}$. Note that the index n allows the operator to propose different prices to different customers or, more realistically, to different groups of customers (e.g., students, seniors, families). In that case, the model includes as many p_i variables as the number of defined groups.

The expected gain obtained from service $i \in \mathcal{C} \setminus \{0\}$ can be derived directly from the demand expression (2.7) and the price specification:

$$G_i = \frac{1}{R} \sum_{n=1}^{N} \sum_{r=1}^{R} p_{in} w_{inr}. \tag{2.38}$$

As the price is an endogenous variable, (2.38) is non-linear. We note that, depending on the context, the price can be modeled in a continuous way (as a real-valued variable) or in a discrete way (with a predetermined set of values). We explore the linearization techniques that need to be developed in both cases in Sections 2.3.1 and 2.3.2.

2.3.1 Continuous price

The product of a binary and a continuous variable can be linearized if an upper bound on the latter is known, which in this case can be set by the operator. Assume that the price p_{in} is bounded between a lower bound $a_{in} \in \mathbb{R}$ and an upper bound $b_{in} \in \mathbb{R}$. We define the variables $\eta_{inr} = p_{in} w_{inr}$ capturing the product of the two, with linearizing constraints (2.39)–(2.42):

$$\begin{aligned}
a_{in} w_{inr} &\leq \eta_{inr}, && \forall i \in \mathcal{C} \setminus \{0\}, n, r, && (2.39)\\
\eta_{inr} &\leq b_{in} w_{inr}, && \forall i \in \mathcal{C} \setminus \{0\}, n, r, && (2.40)\\
p_{in} - (1 - w_{inr}) b_{in} &\leq \eta_{inr}, && \forall i \in \mathcal{C} \setminus \{0\}, n, r, && (2.41)\\
\eta_{inr} &\leq p_{in} - (1 - w_{inr}) a_{in}, && \forall i \in \mathcal{C} \setminus \{0\}, n, r. && (2.42)
\end{aligned}$$

It is easy to verify that constraints (2.39) and (2.40) are binding when $w_{inr} = 0$, and impose $\eta_{inr} = 0$, and constraints (2.41) and (2.42) are binding when $w_{inr} = 1$, and impose $\eta_{inr} = p_{in}$. The expected gain G_i is then obtained by replacing the product of variables $p_{in} w_{inr}$ by the variables η_{inr} in (2.38), i.e.,

$$G_i = \frac{1}{R} \sum_{n=1}^{N} \sum_{r=1}^{R} \eta_{inr}. \qquad (2.43)$$

Motivated by the definition of the η_{inr} variables, we can derive the following valid inequalities:

$$d_{0nr} w_{0nr} + \sum_{j \in \mathcal{C} \setminus \{0\}} (d_{jnr} w_{jnr} + \beta_{jn} \eta_{jnr}) \geq z_{inr}, \qquad \forall i \in \mathcal{C}, n, r, \qquad (2.44)$$

where $d_{inr} = f^d_{in}(x^d_{in}) + \xi_{inr}$ refer to the constant term included in U_{inr} (see (2.8)). These constraints are the linearized version of

$$\sum_{j \in \mathcal{C}} U_{jnr} w_{jnr} \geq z_{inr}, \qquad \forall i \in \mathcal{C}, n, r. \qquad (2.45)$$

Indeed, constraint (2.44) is equivalent to constraints (2.23)-(2.24), which set the choice variable equal to 1 for the service with the highest discounted utility. We have compared the computational times for different instances with and without (2.44), and we observe that it helps to obtain a better LP relaxation, which allows to solve the MILP more efficiently.

It is easy to prove the equivalence with constraints (2.23)-(2.24). Assume that z_{i^*nr} is the largest discounted utility associated with individual n and draw r, i.e., $z_{i^*nr} \geq z_{inr}$, $\forall i \in \mathcal{C}_n, i \neq i^*$. Given that only one of the variables w_{inr} associated with individual n and draw r can be equal to 1 due to constraints (2.20), if $w_{j^*nr} = 1$ for $j^* \neq i^*$, then constraint (2.44) can be written as $U_{j^*nr} \geq z_{jnr}$ for any $j \in \mathcal{C}_n$, and in particular for $j = i^*$. Since $y_{j^*nr} = 1$ (using (2.21)), the valid inequality (2.44) can be written as $z_{j^*nr} \geq z_{i^*nr}$ for service i^*, which implies, together with the initial assumption $z_{i^*nr} \geq z_{j^*nr}$, that $z_{j*nr} = z_{i*nr}$. This is equivalent to (2.23)-(2.24), since $U_{nr} = \max_{i \in C} z_{inr} = z_{j*nr}$.

2.3.2 Discrete price

When operators are not interested in the full spectrum of prices, but in a specific subset (e.g., round prices only), it is more convenient to assume that p_{in} can only take a finite number of predetermined values, called price levels. Note that any integer variable that is bounded and can take only a finite number of values can be written as a linear combination of binary variables. This enables to characterize the price levels with a smaller number of binary variables, as opposed to associating a binary variable with each price level. However, p_{in} is not defined as an integer variable, and neither its price levels. We can express them as integer numbers by setting a precision of k decimals and multiplying them by 10^k.

Consider $p_{in} \in 1/10^k\{a_{in}, \ldots, b_{in}\}$, where $\{a_{in}, \ldots, b_{in}\}$ are the integer price levels for service $i \in \mathcal{C} \setminus \{0\}$ and individual n, sorted from the smallest level (a_{in}) to the largest (b_{in}). We define L_{in} binary variables $\lambda_{in\ell}$ for each service $i \in \mathcal{C} \setminus \{0\}$ and each individual n, where L_{in} is the smallest integer such that $b_{in} - a_{in} \leq 2^{L_{in}} - 1$, i.e., $L_{in} = \lceil \log_2(b_{in} - a_{in} + 1) \rceil$. We can write p_{in} as follows:

$$p_{in} = \frac{1}{10^k}\left(a_{in} + \sum_{\ell=0}^{L_{in}-1} 2^\ell \lambda_{in\ell}\right), \qquad \forall i \in \mathcal{C} \setminus \{0\}, n. \tag{2.46}$$

Notice that (2.46) can generate prices above b_{in} if $\lceil \log_2(b_{in} - a_{in} + 1) \rceil > \log_2(b_{in} - a_{in} + 1)$. If it is important to generate prices below b_{in}, the following constraint must be included:

$$a_{in} + \sum_{\ell=0}^{L_{in}-1} 2^\ell \lambda_{in\ell} \leq b_{in}, \qquad \forall i \in \mathcal{C} \setminus \{0\}, n. \tag{2.47}$$

The expected gain G_i is written as

$$G_i = \frac{1}{R}\frac{1}{10^k}\left(\sum_{n=1}^{N}\sum_{r=1}^{R} a_{in} w_{inr} + \sum_{\ell=0}^{L_{in}-1} 2^\ell \lambda_{in\ell} w_{inr}\right). \tag{2.48}$$

In order to linearize the product of the binary variables $\lambda_{in\ell}$ and w_{inr} in (2.48), we introduce the binary variables $\alpha_{inr\ell} = \lambda_{in\ell} w_{inr}$, so that G_i becomes linear, with linearizing constraints (2.49)–(2.51):

$$\lambda_{in\ell} + w_{inr} \leq 1 + \alpha_{inr\ell}, \qquad \forall i \in \mathcal{C} \setminus \{0\}, n, r, \ell, \tag{2.49}$$

$$\alpha_{inr\ell} \leq \lambda_{in\ell}, \qquad \forall i \in \mathcal{C} \setminus \{0\}, n, r, \ell, \tag{2.50}$$

$$\alpha_{inr\ell} \leq w_{inr}, \qquad \forall i \in \mathcal{C} \setminus \{0\}, n, r, \ell. \tag{2.51}$$

Hence, (2.48) can be linearly expressed as follows:

$$G_i = \frac{1}{R}\frac{1}{10^k}\left(\sum_{n=1}^{N}\sum_{r=1}^{R} a_{in} w_{inr} + \sum_{\ell=0}^{L_{in}-1} 2^\ell \alpha_{inr\ell}\right). \tag{2.52}$$

Furthermore, valid inequality (2.44) can also be derived for discrete prices:

$$d_{0nr} w_{0nr} + \sum_{j \in \mathcal{C}_n \setminus \{0\}} \left(d_{jnr} w_{jnr} + \frac{1}{10^k} \beta_{jn} \sum_{\ell=0}^{L_{jn}-1} 2^\ell \alpha_{jnr\ell} \right) \geq z_{inr}, \qquad \forall i \in \mathcal{C}, n, r, \tag{2.53}$$

where d_{inr} is defined in this case as $d_{inr} = f^d_{in}(x^d_{in}) + \xi_{inr} + \frac{1}{10^k}\beta_{in} a_{in}$, with $\beta_{0n} = 0, \forall n$.

2.3.3 Expected profit

The generated revenues are provided by (2.43) in the case of continuous price and by (2.52) if a discrete price is assumed. Regarding the costs, we assume that the operating cost of service $i \in \mathcal{C} \setminus \{0\}$ is calculated as

$$C_i = \sum_{q=1}^{Q} (f_{iq} + v_{iq} c_{iq}) y_{iq}, \tag{2.54}$$

where f_{iq} is the fixed cost and v_{iq} is the cost per sold unit associated with service i and capacity level c_{iq}.

The expected profit is computed by subtracting the total operating costs from the generated revenues. The resulting objective function is the following:

$$\max \sum_{i \in \mathcal{C} \setminus \{0\}} (G_i - C_i). \tag{2.55}$$

The constraints that are included in the optimization model are itemized next:

- Utility: (2.8)

- Availability: (2.11), (2.12)
- Discounted utility: (2.14), (2.15), (2.16), (2.17)
- Choice: (2.20), (2.21), (2.23), (2.24)
- Capacity allocation: (2.28), (2.29), (2.30) (2.31)
- Pricing: continuous (2.39), (2.40), (2.41) (2.42) or discrete (2.47), (2.49), (2.50) (2.51)
- Valid inequality: (2.44) for continuous prices and (2.53) for discrete prices.

Table A.1 in Appendix A summarizes the main notations used in the model for the reader's convenience, organized by sets, parameters, variables and aggregated quantities. For the sake of simplicity, we do not make the distinction between the terms alternatives and services, as it is done in Sections 2.2 and 2.3, respectively, and we refer to them simply as alternatives.

2.4 Case study

The objective of this case study is twofold. We first deal with the integration of an existing DCM from the literature in the MILP described in Section 2.3, and we then perform several experiments to illustrate the extent of the resulting formulation.

Discrete choice model

The challenge of the first goal consists in embedding a non trivial DCM externally developed in the profit maximization problem. We rely on the case study of a parking services operator, which is motivated by a published DCM for parking choice (Ibeas et al., 2014), whose data was kindly provided by the authors and used to perform the experiments discussed next.

The parking choice model aims at addressing the economic viability of an underground parking in the area of study. In order to adapt the case study to the application described in Section 2.3, we assume a park and ride situation, where the parking facilities have public transportation connections and the users leave their vehicles during the day in order to commute to their final destination with public transportation. In this way, we circumvent the inherent dynamic behavior of general parking facilities, where the spots are occupied and liberated indistinctly, and we define a setting where as soon as the parking spot is taken, it will not be available for the users subsequently arriving next (within the time horizon being evaluated).

We note that this characterization assumes fully informed drivers, i.e., drivers know if there is a remaining parking spot in each parking facility, and where this spot is located, which is not always the case. A more comprehensive analysis of a problem handling parking facilities should include additional features, such as the fact that finding the last free parking space is a considerable time-consuming effort, while other parking spaces can be freed. In any case, the focus here is not on parking management, but on showing that a disaggregate demand model from the literature can be directly accommodated in MILP thanks to the proposed framework.

The choice set consists of three services: paid on-street parking (PSP), paid parking in an underground car park (PUP) and free on-street parking (FSP). Since the latter does not provide any revenue to the operator, it represents the opt-out option, and therefore it is always available to all users. We assume that the parking facilities are either open to everyone or not offered at all, and that the associated price is the same for everyone. Furthermore, we assume that $\mathcal{C}_n = \mathcal{C}$, $\forall n$, i.e., all users consider all facilities when deciding where to park.

Together with the technical variables used for linearization purposes, the model for this case study contains the availability variables y_{in}, which are reduced to y_i (i.e., $y_{in} = y_i, \forall n$) based on the above-mentioned assumption, and y_{inr}, which model the availability at scenario level; the binary variables w_{inr}, which represent the choice; and the price variables p_{in}, which are simplified to p_i as a single price is assumed, and we decide to model it continuously.

The utility specification of the demand model is given by Ibeas et al. (2014). They define a mixture of logit models (see (2.6) for the associated probability expression) to describe the behavior of potential car park users. Since the model has a linear-in-parameter formulation, it can be described with a specification table, which contains as many columns as alternatives in the model (three in this case) plus one column including the values of the coefficients, and as many rows as parameters within the DCM. Table 2.1 contains the estimates of the parameters and the specification associated with each parking service.

Table 2.1: Specification table of the mixtures of logit model in Ibeas et al. (2014)

		FSP	PSP	PUP
$\mathrm{ASC_{PSP}}$	32	0	1	0
$\mathrm{ASC_{PUP}}$	34	0	0	1
β_{AT}	$\sim \mathcal{N}(-0.788, 1.06)$	$\mathrm{AT_{FSP}}$	$\mathrm{AT_{PSP}}$	$\mathrm{AT_{PUP}}$
β_{TD}	-0.612	$\mathrm{TD_{FSP}}$	$\mathrm{TD_{PSP}}$	$\mathrm{TD_{PUP}}$
$\beta_{\mathrm{Origin_{INT\_FSP}}}$	-5.76	$\mathrm{Origin_{INT\_FSP}}$	0	0
β_{FEE}	$\sim \mathcal{N}(-32.3, 14.2)$	0	$\mathrm{FEE_{PSP}}$	$\mathrm{FEE_{PUP}}$
$\beta_{\mathrm{FEE_{PSP(LoIn)}}}$	-11	0	$\mathrm{FEE_{PSP}LoIn}$	0
$\beta_{\mathrm{FEE_{PSP(Res)}}}$	-11.4	0	$\mathrm{FEE_{PSP}Res}$	0
$\beta_{\mathrm{FEE_{PUP(LoIn)}}}$	-13.7	0	0	$\mathrm{FEE_{PUP}LoIn}$
$\beta_{\mathrm{FEE_{PUP(Res)}}}$	-10.7	0	0	$\mathrm{FEE_{PUP}Res}$
$\beta_{\mathrm{AgeVeh}\leq 3}$	4.04	0	0	$\mathrm{AgeVeh} \leq 3$

The random coefficients are the ones associated with the access time to the parking place once the user arrives to the parking area (AT) and the parking fee (FEE), and are denoted by β_{AT} and β_{FEE}, respectively. As mentioned by the authors, the latter is related to an hour of use of the parking place, regardless of the time that the spot was needed. The units are not specified. Both parameters are assumed to be normally distributed and correlated, with $cov(\mathrm{AT}, \mathrm{FEE}) = -12.8$.

The other variables appearing in the utility specification are the following: access time to the destination from the parking spot (TD), an indicator variable that is 1 if the origin of the trip is internal to the town ($\mathrm{Origin}_{\mathrm{INT\_FSP}}$), an indicator variable that is 1 if the income of the user is below 1200/month (LoIn), an indicator variable that is 1 if the user is resident (Res), and an indicator variable that is 1 if the age of the vehicle is lower than 3 years ($\mathrm{AgeVeh}_{\leq 3}$). Two interactions to address the variations in taste among users are considered: FEE with having a low income and FEE with being resident.

MILP model

For the sake of illustration, and to avoid solving huge optimization problems, we define a sample of $N = 50$ users, which are randomly selected among the 197 users available in the provided dataset. Despite the reduction in size, this sample still represents a realistic example to test the formulation on. We define the priority list as the order of the users in the sample, which can be interpreted as a random arrival.

In the optimization model described in Section 2.3, we assume that the price (FEE) is the only endogenous variable. It appears linearly in the utility function (as required), as well as the other demand variables, even though linearity on these is not necessary because they are not decision variables of the MILP model. The values of p_{PSP} ($\mathrm{FEE}_{\mathrm{PSP}}$) and p_{PUP} ($\mathrm{FEE}_{\mathrm{PUP}}$) will be determined by the model, whereas the exogenous demand variables will be replaced by their corresponding values in the data. As a stated preference (SP) survey was conducted, we consider one of the choice situations presented to the respondents in order to characterize the attributes of the alternatives previously described.

We perform several experiments to evaluate different features of the framework. In the first three, we assume a fixed capacity of $c_{\mathrm{PSP}} = c_{\mathrm{PUP}} = 20$ spots, which is large enough to be realistic for the size of the sample but restrictive enough to force some users to opt-out because there is not enough room for everyone. Moreover, the fixed and variable costs of the paid alternatives are set to 0, which turns the objective function into the expected revenue obtained from the offered services. Notice that all services are offered by the operator in this case as we assume that either the service is offered to everyone or not offered at all, so from a revenue maximization point of view it does not make sense not to open a service. The last experiment deals with the profit maximization problem as such. All the experiments have been implemented in C++ using ILOG Concert Technology to access CPLEX 12.8, and all the instances were performed using 12 threads in a 3.33 GHz Intel Xeon X5680 server running a 64-bit Ubuntu 16.04.2.

Section 2.4.1 provides an assessment of the computational time and obtained solutions both for the uncapacitated and capacitated case. Section 2.4.2 deals with the segmentation of individuals by differentiating the price among market segments and by grouping individuals with similar behavior, which involves a reduction of the size of the problem. Section 2.4.3 evaluates the impact of the priority list when a random arrival of users is assumed, both when the users are individually considered and when they are grouped. Finally, Section 2.4.4 analyzes the computational expense and impact on the results of the maximization of the profit subject to a capacity allocation strategy.

2.4.1 Price calibration

In this section, we determine the optimal price of PSP and PUP so that the revenue of the operator is maximized, both with and without capacity restrictions on these services. Based on the values of the variable FEE in the data, we assume $p_{\text{PSP}} \in [0.5, 0.65]$ and $p_{\text{PSP}} \in [0.7, 0.85]$. We analyze the performance of the framework with respect to the number of simulation draws by running 5 replications for each value of R, each replication corresponding to an independent generation of R draws of the error terms $\varepsilon_{in}, \forall i \in \mathcal{C}, n$. Furthermore, we evaluate the obtained solutions for a large number of draws, i.e., we compute the expected optimal revenue with the obtained optimal prices with a large number of draws in order to assess their quality, and we determine the number of draws that will be considered for the following experiments.

Uncapacitated case

We assume first that both PSP and PUP have unlimited capacity, i.e., constraints (2.28)–(2.29) are ignored. We include them back later in order to analyze the expected increase in solution time due to the increase in complexity and the differences with respect to optimal prices and expected demand.

Table 2.2 presents aggregate statistics of the expected optimal revenue and computational time. We observe that as R increases, the standard deviation decreases, which shows the stability of the obtained results as the number of draws becomes larger. Regarding the computational time, we observe the expected exponential growing with respect to R, which becomes particularly noticeable from $R = 100$ (average computational time of 24 minutes) to $R = 250$ (average computational time of 1.76 hours).

Figures 2.2 and 2.3 provide the boxplots for the optimal prices of the paid alternatives and the expected demand for all alternatives, respectively. In all cases, the interquartile range (the difference between the upper and lower quartiles) decreases with respect to R,

Table 2.2: Computational results of the revenue maximization problem for the uncapacitated cases for 5 replications (independent generation of draws)

R	Expected optimal revenue				Computational time (min)		
	Min.	Avg.	Max.	Std. dev.	Min.	Avg.	Max.
2	25.794	26.790	28.284	1.030	0.002	0.003	0.003
5	26.574	27.258	27.906	0.565	0.007	0.009	0.013
10	26.547	27.262	27.642	0.425	0.042	0.050	0.060
25	26.653	26.947	27.087	0.170	0.159	0.307	0.464
50	26.749	26.867	27.071	0.126	1.364	3.267	4.844
100	26.787	26.872	27.059	0.111	15.195	24.173	32.159
250	26.792	26.889	26.967	0.077	55.721	105.445	198.026

which shows a decrease in the variability of the obtained results as the number of draws increases.

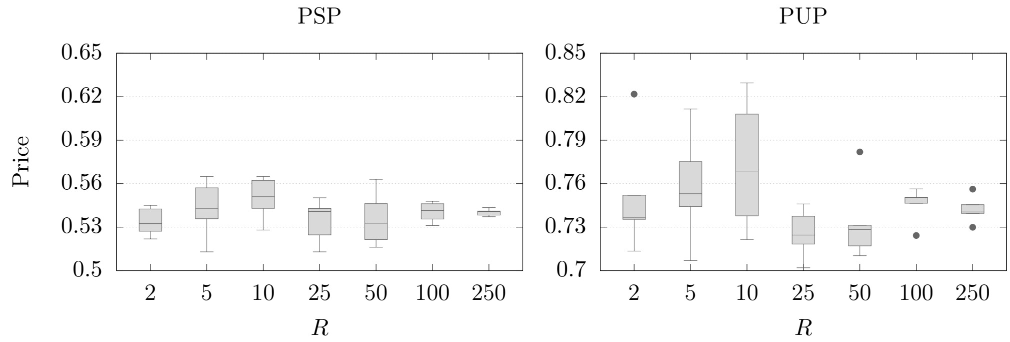

Figure 2.2: Boxplot of the optimal prices for the uncapacitated case for 5 replications (independent generation of draws)

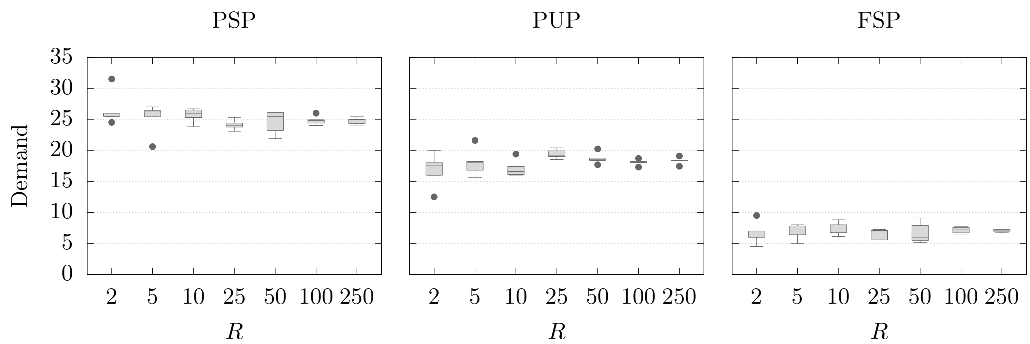

Figure 2.3: Boxplot of the expected demand for the uncapacitated case for 5 replications (independent generation of draws)

Capacitated case

Table 2.3 shows the increase in computational time with respect to the uncapacitated case, which is expected since the implementation of the priority list and the tracking of the occupancy for each alternative hugely complicate the solution approach. Similarly to the uncapacitated case, we can observe in Figures 2.4 and 2.5 a decreasing variability of the optimal prices and expected demand as R increases.

Table 2.3: Computational results of the revenue maximization problem for the capacitated case for 5 replications (independent generation of draws)

R	Expected optimal revenue				Computational time (min)		
	Min.	Avg.	Max.	Std. dev.	Min.	Avg.	Max.
2	24.972	26.183	27.699	1.011	0.006	0.010	0.022
5	25.706	26.246	26.779	0.504	0.093	0.115	0.142
10	25.616	26.404	26.810	0.486	1.157	1.656	2.155
25	25.738	26.029	26.297	0.205	7.714	13.160	22.793
50	25.676	26.000	26.280	0.244	38.119	59.885	69.313
100	25.835	26.015	26.131	0.109	204.940	300.442	450.428
250	25.933	25.977	26.067	0.053	657.679	1261.398	2020.310

Both PSP and PUP are more expensive in the capacitated case. Since the demand of PSP was already higher than its current capacity in the uncapacitated case, its price can be increased so that the operator obtains a higher revenue from the users accessing the service. In the case of PUP, the price is also higher, but the demand is similar to the one in the uncapacitated case, which might also be influenced by the capacity restriction on PSP, since normally the opt-out option is the least attractive alternative. However, FSP is experiencing an increase in its demand because it is capturing the users that cannot be allocated due to capacity limitations or that are not willing to pay the current price of the paid alternatives.

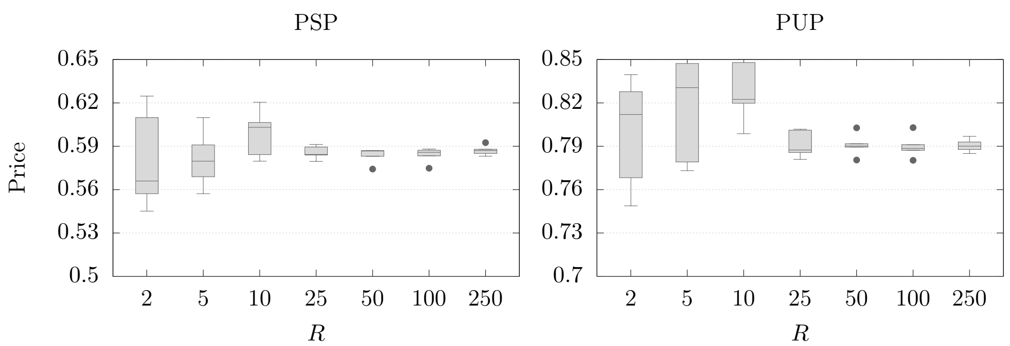

Figure 2.4: Boxplot of the optimal prices for the capacitated case for 5 replications (independent generation of draws)

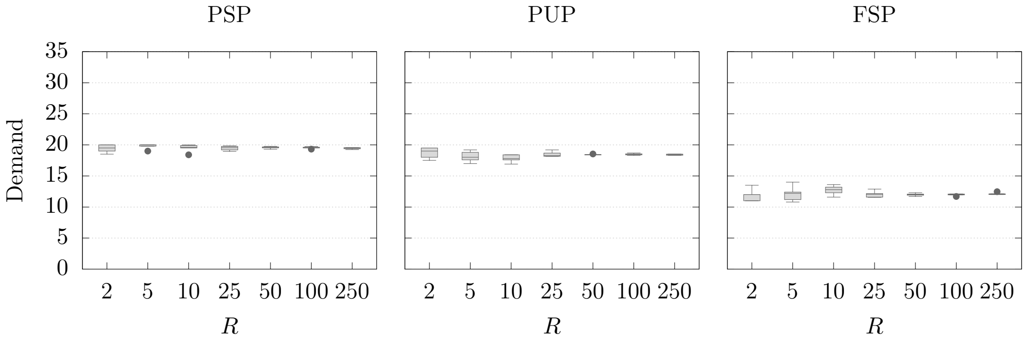

Figure 2.5: Boxplot of the expected demand for the capacitated case for 5 replications (independent generation of draws)

Evaluation of optimal solutions

We now compare the expected optimal revenues for the different tested values of R and 5 replications with the expected optimal revenues obtained by evaluating the prices of PSP and PUP of the MILP solution with a large number of draws, namely $R^* = 10^6$. Table 2.4 includes the relative differences (in %) with the expected optimal revenues obtained by evaluating the optimal prices as reference values. Such differences decrease as R increases, which indicate that the expected gain becomes more robust with larger values of R. In this experiment, when $R \geq 25$, the relative difference in both cases is lower than 1%.

Table 2.4: Relative differences of the expected optimal revenues obtained by solving the MILP model with respect to the expected optimal revenues obtained by evaluating the optimal prices with $R^* = 10^6$ for 5 replications (independent generation of draws)

R	Relative diff. (%) (uncapacitated)			Relative diff. (%) (capacitated)		
	Minimum	Average	Maximum	Minimum	Average	Maximum
2	0.803	2.972	5.214	0.060	3.258	8.139
5	0.048	2.048	4.254	3.377	3.856	4.252
10	1.132	2.217	3.461	0.937	2.952	4.550
25	0.376	0.570	0.898	0.131	0.707	1.424
50	0.077	0.356	0.732	0.167	0.727	1.320
100	0.057	0.302	0.688	0.192	0.443	0.849
250	0.151	0.249	0.360	0.020	0.171	0.503

Required number of draws

The required number of draws is closely related to the variance of the error terms of the DCM. That is, if the DCM is highly deterministic (low variance of the error terms), the choices will mostly be driven by the systematic component of the utility function (V_{in}), and a low number of draws will be required. We illustrate the level of stochasticity in this case by considering an error rate defined as the ratio of choices estimated by the model that do not match the deterministic choice (the choice obtained in the absence of error component) to the total number of estimated choices. Notice that we need to set the distributed parameters β_{AT} and β_{FEE} equal to their means in order to generate the deterministic choices. The error rate fluctuates between 50% and 52% for the different tested values of R in the capacitated case, which indicates that the level of stochasticity of the mixed logit model is considerable.

Nevertheless, as seen in the different analyses, the obtained quantities are stable enough for $R \geq 25$, in the sense that a low variability is experienced. For the sake of illustration, we consider $R = 50$ for the experiments developed in Sections 2.4.2 and 2.4.3, as it provides a good compromise between computational time and precision of the results. Since the problem containing the capacity as decision variable is computationally more expensive, Section 2.4.4 considers $R = 25$.

2.4.2 Population segmentation

The disaggregate representation of the demand provides a great deal of flexibility when it comes to integrate population segmentation strategies within the framework. In this section, we test a price differentiation scheme for two market segments based on the residency in the area of interest, and a grouping of individuals with similar behavior. In the first experiment, the users are still modeled individually, whereas in the second experiment the consideration of user groups reduces the size of the formulation as a single individual represents all the members within the corresponding group. In both cases, capacity restrictions are assumed and 5 replications (independent generation of draws) for each value of R are run.

Price differentiation

Imagine that the municipality provides reduced fees to residents who want to access one of the paid alternatives. This is actually done in many cities, where residents get reduced prices for common parking services or even exclusive areas, where only them have the right to park. In this case, we assume a discount factor that is applied to the prices offered to non residents.

Regarding the operator's revenue, two situations are considered: (1) the difference between the actual price of the service and the contribution of the resident is paid by the municipality in the form of a subsidy and, therefore, contributes to the revenue of the operator, and (2) the operator is obliged by the municipality to offer reduced fees to residents, without any other compensation than the right to operate the parking. In both situations, the reduced prices have an impact on the utility functions of the residents, and consequently on their choice. In situation 1, the revenue is not impacted whereas in situation 2 the reduced fares cause a decrease in the total gain.

Since residents only pay a part of the fee that non residents pay, the former users might be attracted to higher fares, and we therefore expect the prices of both services to increase. We modify the price bounds of PSP and PUP as follows: $p_{\text{PSP}} \in [0.6, 1.2]$ and $p_{\text{PUP}} \in [0.8, 1.4]$. The average expected optimal revenue as well as the interval defined by the lowest and highest obtained values for the optimal prices and expected demand for 5 replications are included in Tables 2.5 and 2.6 for situations 1 and 2, respectively. In both situations, the higher the discount, the higher the prices being offered, as expected. However, this increase is more moderate in situation 2 because it leads to a decrease in the total gain.

Table 2.5: Average revenue and range of obtained results for 5 replications (independent generation of draws) for different subsidies in situation 1 for $R = 50$ (R denotes resident and NR denotes non resident)

Disc. (%)	Revenue	Prices (NR)		Demand (PSP)		Demand (PUP)		Demand (FSP)	
		PSP	PUP	R	NR	R	NR	R	NR
20	28.94	[0.63,0.66]	[0.86,0.89]	[10.1,11.0]	[8.7,9.3]	[7.0,7.9]	[10.7,11.4]	[3.5,4.1]	[7.7,8.2]
25	29.74	[0.67,0.70]	[0.89,0.94]	[10.4,11.1]	[7.8,8.7]	[7.5,8.1]	[10.0,10.9]	[3.1,3.7]	[8.7,9.7]
30	30.57	[0.69,0.71]	[0.91,0.95]	[10.8,11.5]	[7.7,8.4]	[7.9,8.7]	[9.8,10.4]	[2.2,2.8]	[9.5,9.9]
40	32.12	[0.76,0.77]	[0.99,1.05]	[11.3,12.8]	[6.1,6.8]	[7.8,9.0]	[8.7,9.4]	[1.3,1.7]	[12.1,13.0]
50	33.91	[0.84,0.95]	[1.14,1.24]	[10.9,12.6]	[2.6,5.6]	[8.6,9.5]	[6.7,8.1]	[0.7,1.7]	[15.0,17.4]

Table 2.6: Average revenue and range of obtained results for 5 replications (independent generation of draws) for different subsidies in situation 2 for $R = 50$ (R denotes resident and NR denotes non resident)

Disc. (%)	Revenue	Prices (NR)		Demand (PSP)		Demand (PUP)		Demand (FSP)	
		PSP	PUP	R	NR	R	NR	R	NR
20	26.26	[0.63,0.66]	[0.86,0.89]	[10.1,11.0]	[8.7,9.3]	[7.0,7.9]	[10.7,11.4]	[3.5,4.1]	[7.7,8.2]
25	26.13	[0.67,0.69]	[0.90,0.93]	[10.4,11.4]	[7.8,8.7]	[7.1,8.1]	[10.3,10.7]	[3.3,3.6]	[8.7,9.7]
30	25.93	[0.69,0.71]	[0.91,0.94]	[10.8,11.5]	[7.7,8.6]	[7.9,8.7]	[10.2,10.5]	[2.2,2.8]	[9.2,9.9]
40	25.08	[0.71,0.76]	[0.96,0.99]	[11.3,12.2]	[6.1,8.2]	[8.7,9.6]	[9.2,9.8]	[1.0,1.4]	[10.5,12.4]
50	23.77	[0.71,0.77]	[1.01,1.05]	[11.4,12.3]	[6.7,8.4]	[9.4,10.3]	[7.9,9.0]	[0.2,0.6]	[11.1,12.6]

In terms of the expected demand, we observe that the higher the discount, the lower the expected non resident demand of PSP and PUP and the higher the number of non residents deciding to opt-out, as they are not willing to pay the offered fees and choose FSP instead. Certainly, the expected resident demand experiences the opposite, since it decreases for FSP and increases for PSP and PUP as the discount increases. Note that a 20% discount almost has no impact on the optimal prices and expected demand, i.e., a larger discount needs to be put in place in order to alter the choices of the individuals.

Grouping of individuals

In this experiment, we define $\mathcal{G}$ groups of individuals with homogeneous behavior. We keep the same notation and denote the groups by n and their size (number of individuals) by θ_n. We assume that $\theta_n \leq c_i, \forall i \in \mathcal{C} \setminus \{0\}, n$, i.e., the size of the groups does not exceed the capacity of the alternatives.

We rely on the socioeconomic variables that are present in the DCM to define groups with homogeneous behavior. More precisely, we allow for all possible combinations of the values of the binary variables Res (residency), $\text{Origin}_{\text{INT\_FSP}}$ (origin of the trip), LoIn (low income level) and $\text{AgeVeh}_{\leq 3}$ (age of the vehicle). Table 2.7 includes the user groups derived from such combinations and the sizes within the considered sample ($N = 50$). Notice that 4 of the groups are not represented in the sample, so $\mathcal{G} = 12$.

Table 2.7: Definition of user groups

n	θ_n	Res	$\text{Origin}_{\text{INT\_FSP}}$	LoIn	$\text{AgeVeh}_{\leq 3}$	n	θ_n	Res	$\text{Origin}_{\text{INT\_FSP}}$	LoIn	$\text{AgeVeh}_{\leq 3}$
1	2	1	1	1	1	9	0	0	1	1	1
2	10	1	1	1	0	10	0	0	1	1	0
3	1	1	1	0	1	11	1	0	1	0	1
4	2	1	1	0	0	12	0	0	1	0	0
5	6	1	0	1	1	13	7	0	0	1	1
6	1	1	0	1	0	14	11	0	0	1	0
7	1	1	0	0	1	15	5	0	0	0	1
8	0	1	0	0	0	16	4	0	0	0	0

The values of the remaining variables present in the DCM (the attributes of the alternatives) are the same across individuals because they correspond to one of the choice situations of the SP experiment performed in Ibeas et al. (2014). This might not be the case in other contexts (e.g., revealed preference data). If so, each variable needs to be set to a single value representing the individuals in the group (e.g., the average).

Concerning the formulation, we need to make some assumptions on the access of groups to the services with respect to capacity restrictions. In order to remain as consistent as possible to the individual-based model, we assume that the groups cannot be split, i.e., the group does not have access to an alternative if there is not enough capacity to accommodate all its individuals. Furthermore, the priority list needs to be respected, which is defined by the numbering introduced in Table 2.7, but we allow group $n+1$ to have access to a service provided that it fits even if group n could not access because its size exceeded the remaining capacity.

Hence, we only need to replace constraints (2.28)–(2.29) with their adaptation at the group level given by constraints (2.56)–(2.57). Constraints (2.56) become active when there is no room for group n, i.e., $\sum_{m=1}^{n-1} \theta_m w_{imr} > c_i - \theta_n$, as it forces y_{inr} to be equal to 0. Similarly, when there is room for group n, i.e., $\sum_{m=1}^{n-1} \theta_m w_{imr} \leq c_i - \theta_n$, constraints

(2.57) implies $y_{inr} = 1$. Notice that there is always room for group $n = 1$ (first group in the priority list) thanks to the assumption $\theta_n \leq c_i, \forall i \in \mathcal{C} \setminus \{0\}, n$.

$$\sum_{m=1}^{n-1} \theta_m w_{imr} \leq (c_i - \theta_n) y_{inr} + \left(\sum_{m=1}^{n-1} \theta_m \right) (1 - y_{inr}), \quad \forall i \in \mathcal{C}, n > 1, r, \tag{2.56}$$

$$(c_i - \theta_n + 1)(y_{in} - y_{inr}) \leq \sum_{m=1}^{n-1} \theta_m w_{imr}, \quad \forall i \in \mathcal{C}, n > 1, r. \tag{2.57}$$

Table 2.8 presents the aggregate statistics of the expected optimal revenue and computational time for the capacitated case with the user groups defined in Table 2.7. The computational times for the different values of R are much lower than the ones obtained at the individual level in the capacitated case (see Table 2.3), which also suggests an exponential growing with respect to the number of individuals. For example, the average computational time for $R = 250$ is drastically reduced from 21 hours to less than 3 min. We observe that the expected optimal revenues fluctuate within lower values than in the individual case, and the associated standard deviation, though it decreases as R increases, is larger.

Table 2.8: Computational results for the user groups defined in Table 2.7 for 5 replications (independent generation of draws)

R	Expected optimal revenue				Computational time (min)		
	Min.	Avg.	Max.	Std. dev.	Min.	Avg.	Max.
2	22.950	26.577	29.275	2.501	0.000	0.001	0.001
5	23.314	24.822	25.809	0.995	0.002	0.003	0.003
10	24.139	24.524	25.060	0.355	0.004	0.006	0.008
25	23.446	23.884	24.596	0.477	0.023	0.032	0.051
50	23.299	23.670	24.447	0.473	0.093	0.111	0.143
100	23.328	23.608	24.225	0.366	0.357	0.413	0.479
250	23.493	23.662	23.912	0.158	1.846	2.182	2.828

The decrease on the obtained revenues can be explained by the condition placed on the groups of individuals to access a service. Figures 2.6 and 2.7 exhibit that the expected demand for both PSP and PUP is lower, and even though the associated prices are higher, it is not enough to reach the gains obtained at the individual level. In both cases, the interquartile range is wider, even for $R = 250$. The increase in prices is again due to the fact that the number of individuals that want to access the paid services is larger than the available capacity, although in this case it is not reached because of the group access condition.

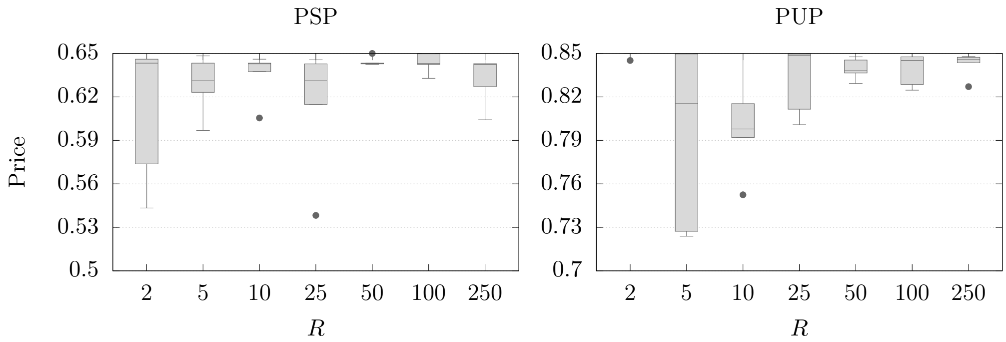

Figure 2.6: Optimal prices for the user groups described in Section 2.4.2 for for 5 replications (independent generation of draws)

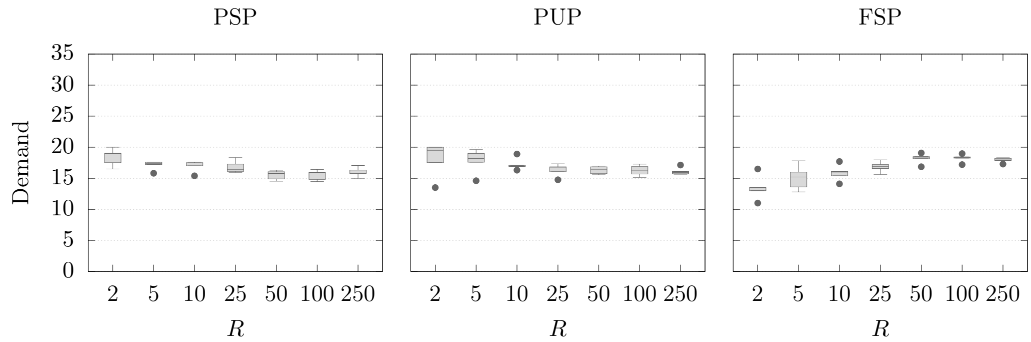

Figure 2.7: Expected demand for the user groups described in Section 2.4.2 for 5 replications (independent generation of draws)

We expect the random ordering of the groups to have a higher impact on the expected optimal revenue than the random ordering at the individual level because of the aggregation of individuals in groups and the fact that the resulting groups are heterogeneous in size. We analyze such effect in Section 2.4.3.

2.4.3 Impact of the arrival of individuals

The priority list described in Section 2.2.2 states the order in which individuals are considered to access the services. As mentioned earlier, the priority list in this case study is defined at random. We can analyze the impact of such priority list on the obtained results by evaluating different priority lists generated at random, both at the individual level and at the group level.

For this experiment we consider situation 2 in Section 2.4.2, i.e., the operator is forced by law to offer a discount on the fares for residents. This allows us to test whether the distribution of residents and non residents within the priority list has a remarkable impact on the expected optimal revenue. For the sake of simplicity, we only allow for the 30% discount rate. We construct 100 different priority lists by shuffling the individuals, so that they might arrive in a different order. We run these instances with $R = 50$, and we assume the same price bounds as in the previous experiment: $p_{\text{PSP}} \in [0.6, 1.2]$ and $p_{\text{PUP}} \in [0.8, 1.4]$.

Figures 2.8 and 2.9 present the histograms of the expected optimal revenue when the users are individually and jointly modeled, respectively. As anticipated in Section 2.4.2, the expected optimal revenue at the group level is more dispersed, with values that fluctuate between 22 and 26 (standard deviation is equal to 0.79), whereas it ranges from 25 to 27 (standard deviation is equal to 0.30) at the individual level. This shows that the distribution of residents and non residents at the group level has a stronger impact on the expected optimal revenue. Similarly to the previous experiment (see Table 2.8), lower values for the expected optimal revenue are obtained.

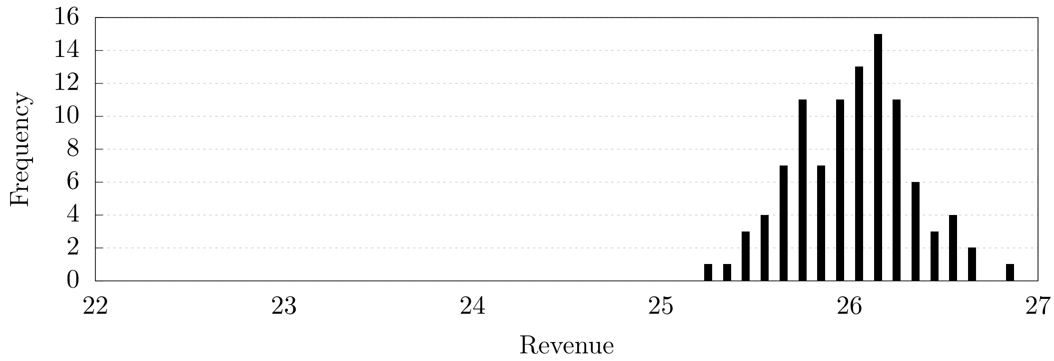

Figure 2.8: Histogram of the expected optimal revenue in situation 2 (discount rate of 30%) for 100 priority lists and $R = 50$ (individual level)

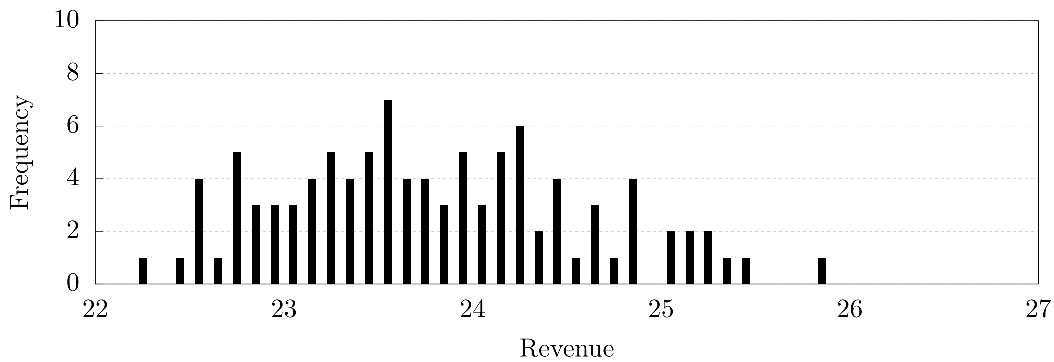

Figure 2.9: Histogram of the expected optimal revenue in situation 2 (discount rate of 30%) for 100 priority lists and $R = 50$ (group level)

We observe a higher concentration of values around an expected optimal revenue of about 26 at the individual level and around 23.5 at the group level. The former is in line with the average value 25.93 obtained in Section 2.4.2 (see Table 2.6). For the latter, and given that a 30% discount in situation 2 does not have a huge impact on the expected optimal revenue at the individual level, we observe that is also in line with the average value 23.67 obtained in Section 2.4.2 (see Table 2.8). This is consistent with the findings of Binder et al. (2017), who show that the aggregate indicators are stable across realizations of a random priority list.

It is worth noticing that the average computational time of the individual approach for a 30% discount in situation 2 (approximately 2 hours) is much larger than its grouping counterpart (approximately 1 min). Hence, one way to address the dispersion manifested at the group level consists in increasing the number of draws while keeping an advantageous computational time. Figure 2.10 shows the histogram at the group level for $R = 100$ (average computational time of 6 min). We still observe some dispersion of the expected optimal revenue but it is less pronounced than for $R = 50$. Furthermore, we note that additional assumptions such as the split of groups could be implemented to diminish the spread and to address the fact that lower values for the expected optimal revenue are obtained.

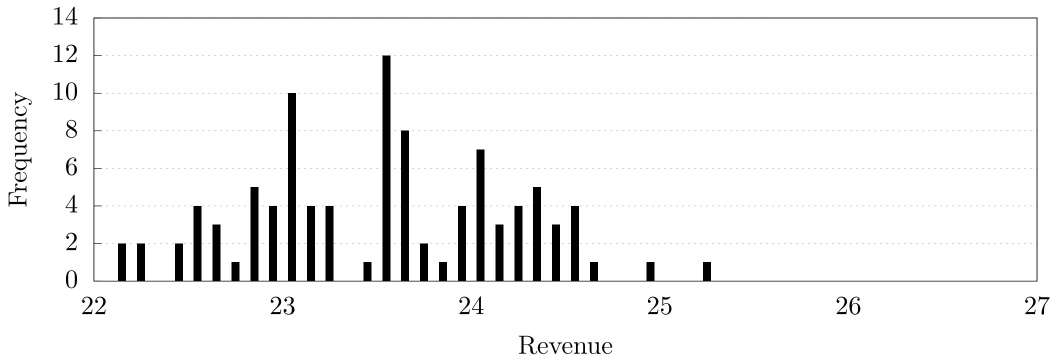

Figure 2.10: Histogram of the expected optimal revenue in situation 2 (discount rate of 30%) for 100 priority lists and $R = 100$ (group level)

2.4.4 Profit maximization through capacity allocation

In this experiment, we test four different capacity levels for PSP and PUP. As described in Section 2.2.3, we replicate the services as many times as capacity levels we want to evaluate. We consider 5, 10, 15 and 20 parking spots ($Q = 4$) for both services, which makes four copies of PSP and four of PUP, each of them with the same utility function but a different capacity level. Together with FSP, this experiment contains 9 different "services."

Note that constraint (2.30) does not force the opening of both paid services, since it might be more convenient from a profit maximization point of view to allocate all the operator's resources to only one of the parking facilities. If we want to make sure that both PSP and PUP are offered, we can replace this constraint by:

$$\sum_{q=1}^{Q} y_{iq} = 1, \quad \forall i \in \mathcal{C} \setminus \{0\}. \tag{2.58}$$

As mentioned in Section 2.3, the cost associated with operating a parking facility is composed of a fixed cost and a variable cost (in this case, a cost per parking spot). We assume that both types of cost are the same among capacity levels. More precisely, $f_{\text{PSP},q} = 1.5$, $v_{\text{PSP},q} = 0.35$, $f_{\text{PUP},q} = 3$ and $v_{\text{PUP},q} = 0.5$ $\forall q$. We also set the price bounds to $p_{\text{PSP}} \in [0.6, 1.2]$ and $p_{\text{PUP}} \in [0.8, 1.4]$ as in Sections 2.4.2 and 2.4.3.

The results for both approaches (considering constraints (2.30) and (2.58), respectively, and running 5 replications for each of them) are included in Table 2.9. We see that it is beneficial to close PUP and only open PSP with the highest level of capacity. Indeed, when we impose that both facilities have to be opened, PUP is offered with a low capacity level (10 in all tested resplications), and a lower profit is generated. The average solution time gives us an idea of the increase in complexity with respect to the revenue maximization problem with fixed capacity. For $R = 25$, it goes from approximately 1 hour to almost 10 hours with constraint (2.30) and more than 11 hours with constraint (2.58).

Table 2.9: Results for the profit maximization with capacities as decision variables for 5 replications (independent generation of draws) for $R = 25$ (in brackets the number of replications for which the associated capacity level was obtained)

	Avg. time (h)	Optimal capacity		Expected demand			Optimal prices		Avg. profit
		PSP	PUP	PSP	PUP	FSP	PSP	PUP	
(2.30)	9.72	20 (5)	NA	[18.2,19.56]	NA	[30.4,31.9]	[0.78,0.85]	NA	6.94
(2.58)	11.4	20 (4)	10 (5)	[14.8,20.0]	[8.76,9.4]	[20.9,26.2]	[0.65,0.73]	[1.00,1.09]	5.99

2.5 Concluding remarks

We have proposed a mixed-integer linear formulation of DCM that is designed to be included in MILP in order to provide a disaggregate demand representation that captures the interactions between the decisions to be optimized by the operator and the individuals. It is general in the sense that it is not limited to simple DCM and it can be embedded in any MILP formulation. The stochasticity of the model is captured by drawing from the distribution of the involved random variables. This enables to avoid the explicit formulation of the choice probabilities and to work directly with the utility functions, using the first principles of utility maximization and SAA.

Concerning the operator, an illustrative MILP model on the profit maximization problem is characterized. The formulation accounts for the preferences of individuals when deciding on the prices via the utility functions. Since the individuals are not captive, i.e., they can leave the market by choosing the opt-out option, there exists a trade-off between the price and their choices.

The results of the case study show that this methodology allows to configure the features of a system (e.g., the price) based on the heterogeneous behavior of individuals. Additional modeling strategies, such as the gathering of individuals in groups of homogeneous behavior, have also been explored. Despite the clear advantages with respect to computational time, we note that the price to pay for this simplification is the fact that the realism of the grouping assumption decreases as the size of the group increases. We also want to notice that for this case study, the obtained results remain quite stable with respect to the number of draws, even for relatively small values thereof. As mentioned in Section 2.4.1, the desired number of draws depends on the accuracy of the DCM. However, we have seen that even with a considerable level of stochasticity associated with the DCM, we can rely on a relatively low number of draws to obtain robust estimates of the quantities of interest.

The disaggregate representation of individuals' preferences, together with the linear nature of the formulation, results in a high-dimensional problem, and therefore solving it is computationally expensive. This is an issue that needs to be addressed because in practice, populations are large and a high number of draws is desirable to be as close as possible to the true value. One possibility is to add valid inequalities to the MILP model in order to be able to solve larger instances, but we notice that they are very likely to depend on the case study, which hinders a universal definition of such inequalities.

Decomposition techniques are convenient in this case to speed up the solution approach, and they represent an alternative to valid inequalities because they can be applied in a general way (although both could also be combined). A Lagrangian decomposition scheme for the choice-based optimization framework is developed in Chapter 3.

3

A Lagrangian decomposition scheme for the choice-based optimization framework

Preliminary ideas related to this chapter are included in the conference paper

Pacheco, M., Lurkin, V., Gendron, B., Sharif Azadeh, S., and Bierlaire, M. (2018). Lagrangian relaxation for the demand-based benefit maximization problem. In *18th Swiss Transport Research Conference (STRC), Ascona, Switzerland*

The work in this chapter has been performed by the candidate under the supervision of Prof. Shadi Sharif Azadeh and Prof. Michel Bierlaire and the collaboration of Prof. Bernard Gendron and Prof. Virginie Lurkin.

3.1 Introduction

In Chapter 2, we introduced a mixed-integer linear formulation of a discrete choice model (DCM) that allows to incorporate a disaggregate demand representation that captures the supply-demand interplay in mixed-integer linear problems (MILP). The disaggregate nature of DCM, together with the associated simulation-based linearization, comes with a high computational complexity, as illustrated in Section 2.4, with large solving times for medium-size instances, hindering its tractability and potential scaling-up. The objective of this chapter is to speed up the solution approach and to provide near-optimal solutions thanks to decomposition techniques, which are procedures applied to optimization problems whose decomposable structure can be advantageously exploited.

By design, the choice-based optimization framework is built on two dimensions that can be addressed separately: the individuals, which represent the most fundamental unit of demand, and the simulation draws. Indeed, each individual n aims at choosing the alternative among the available ones maximizing their utility, and each draw r can be understood as an independent behavioral scenario. Nevertheless, there are some links in the framework that prevent a direct decomposition along these dimensions.

The individuals are tied together through the constraints that keep track of the occupancy of each alternative and ensure that the corresponding capacity is not exceeded. The draws are coupled in the calculation of the total expected demand, which is used to derive other quantities, such as the expected revenue (see Section 2.3). In the following, we rely on this application (revenue maximization problem) to characterize the evaluated decomposition techniques.

In this chapter, we develop an approach based on Lagrangian decomposition in order to rapidly generate upper bounds on the optimal value of the objective function. At the same time, feasible solutions that are close enough to the optimal solutions are obtained. Such solutions represent lower bounds on the optimal value of the objective function, as it is to be maximized. We induce decomposition by creating copies of some of the variables of the problem and dualizing the constraints imposing that the copies should be identical. The multipliers associated with the dualized constraints are then updated on an iterative basis with the well-known subgradient method.

Decomposition methods, and in particular Lagrangian decomposition, are general techniques that need to be tailored to the problem of interest. As the same problem can be seen from different perspectives, there are typically different possibilities when it comes to apply the techniques that are available in the literature (see Section 3.3 for further details). Due to the interrelations between some of the variables and constraints in the choice-based optimization framework, the implementation of some of the reviewed strategies turned out to be unsuitable to our particular case. More precisely, the first tested scheme, which prompts separability for each of the agents involved in the optimization problem (the individuals and the operator), yielded trivial subproblems. Additionally, the precursor of the Lagrangian decomposition framework here proposed might generate irrelevant solutions under some circumstances.

Consequently, one of the contributions of this chapter is the assessment of these decomposition strategies with respect to the choice-based optimization framework. Second, we propose a Lagrangian decomposition scheme that allows for these findings. It comprises the decomposition strategy for obtaining upper bounds and a heuristic that uses the Lagrangian solution to derive feasible solutions at each iteration of the subgradient method. This approach is characterized for the revenue maximization problem and is finally tested on the parking case study introduced in Section 2.4 by scaling up the instances investigated there.

The remaining of the chapter is organized as follows. Section 3.2 provides the different formulations that are considered of the revenue maximization problem upon which the decomposition techniques will be developed. Section 3.3 presents an analysis of the related literature and the considerations that were taken into account during the characterization of the decomposition methods. Section 3.4 introduces the evaluated decomposition strategies and reviews their limitations with respect to the problem under consideration. Section 3.5 describes the Lagrangian decomposition scheme, and Section 3.6 reports the results of the computational experiments. Lastly, Section 3.7 gives some concluding remarks.

3.2 Choice-based revenue maximization problem

As mentioned in Section 3.1, we rely on the revenue maximization problem to outline the decomposition strategies considered in this chapter. In this section, we perform some specifications with respect to the general modeling framework introduced in Sections 2.2 and 2.3 in order to detail the formulation upon which the decomposition scheme in Section 3.5 will be derived.

We assume that the capacity of each service is given, and therefore it is not a decision variable of the problem. Moreover, we assume that all services are made available to all individuals by the operator, as is the case in Section 2.4, which means that the availability variables y_{in} are equal to 1 if service i is considered by customer n, and 0 otherwise. We decide to model the price in a continuous manner.

The revenue maximization problem is depicted in Model 3.1. A summary of the used notations can be found in Appendix A. The objective function (3.1) represents the expected revenue obtained from all services but the opt-out option and all customers. Constraints (3.2) formulate the utility specification, constraints (3.3)–(3.4) provide the linearization of the discounted utility variables, constraints (3.7)–(3.8) associate the choice with the service with the highest discounted utility, constraints (3.9) impose that only one service is chosen per customer and scenario, constraints (3.10) prevent an unavailable service to be chosen, constraints (3.11)–(3.12) ensure that the capacity is not exceeded, constraints (3.13)–(3.16) handle the linearization of the product of the price and the choice variables, and constraints (3.17) provide valid inequalities that help to solve the problem more efficiently (see Sections 2.2.2 and 2.3.1 for further details).

Given the complexity introduced by the tracking of the remaining capacity of the services (see the experiments performed in Section 2.4.1), we rely on the uncapacitated version of the revenue maximization problem as a starting point for the characterization of the decomposition strategies described in Section 3.4. When an unlimited capacity is assumed, constraints (3.11)–(3.12) are pulled out of the model and some other aspects of the formulation are notably simplified. The availability variables at scenario level (y_{inr}) are no longer needed. In the absence of such variables, we can dispense with the discounted utility variables z_{inr} and the associated linearizing constraints. This enables to model the choice directly in terms of the utility variables U_{inr}.

$$
\begin{aligned}
Z = \max \quad & \frac{1}{R}\sum_{i\in\mathcal{C}\setminus\{0\}}\sum_{n=1}^{N}\sum_{r=1}^{R}\eta_{inr} && && (3.1)\\
\text{s.t.} \quad & U_{inr} = \beta_{in}p_{in} + d_{inr} && \forall i \in \mathcal{C}_n, n, r && (3.2)\\
& \ell_{nr} \leq z_{inr} && \forall i \in \mathcal{C}, n, r && (3.3)\\
& z_{inr} \leq \ell_{nr} + M_{inr}y_{inr} && \forall i \in \mathcal{C}, n, r && (3.4)\\
& U_{inr} - M_{inr}(1-y_{inr}) \leq z_{inr} && \forall i \in \mathcal{C}, n, r && (3.5)\\
& z_{inr} \leq U_{inr} && \forall i \in \mathcal{C}, n, r && (3.6)\\
& z_{inr} \leq U_{nr} && \forall i \in \mathcal{C}, n, r && (3.7)\\
& U_{nr} \leq z_{inr} + M_{inr}(1-w_{inr}) && \forall i \in \mathcal{C}, n, r && (3.8)\\
& \sum_{i\in\mathcal{C}} w_{inr} = 1 && \forall n, r && (3.9)\\
& w_{inr} \leq y_{inr} && \forall i \in \mathcal{C}, n, r && (3.10)\\
& \sum_{m=1}^{n} w_{imr} \leq (c_i - 1)y_{inr} + (n-1)(1-y_{inr}) && \forall i \in \mathcal{C}, n > c_i, r && (3.11)\\
& c_i(1-y_{inr}) \leq \sum_{m=1}^{n} w_{imr} && \forall i \in \mathcal{C}, n > 1, r && (3.12)\\
& a_{in}w_{inr} \leq \eta_{inr} && \forall i \in \mathcal{C}\setminus\{0\}, n, r && (3.13)\\
& \eta_{inr} \leq b_{in}w_{inr} && \forall i \in \mathcal{C}\setminus\{0\}, n, r && (3.14)\\
& p_{in} - (1-w_{inr})b_{in} \leq \eta_{inr} && \forall i \in \mathcal{C}\setminus\{0\}, n, r && (3.15)\\
& \eta_{inr} \leq p_{in} - (1-w_{inr})a_{in} && \forall i \in \mathcal{C}\setminus\{0\}, n, r && (3.16)\\
& z_{inr} \leq d_{0nr}w_{0nr} + \sum_{j\in\mathcal{C}\setminus\{0\}} (\beta_{jn}\eta_{jnr} + d_{jnr}w_{jnr}) && \forall i \in \mathcal{C}, n, r && (3.17)\\
& y_{inr}, w_{inr} \in \{0,1\} && \forall i \in \mathcal{C}, n, r && (3.18)
\end{aligned}
$$

Model 3.1: Revenue maximization problem

Model 3.2 presents the uncapacitated version of the revenue maximization problem. Constraints (3.20) specify the utility variables, constraints (3.21)–(3.22) linearize the variable capturing the highest utility through the choice variables, constraints (3.23) impose exactly one service to be chosen per customer and scenario, constraints (3.24)–(3.27) take care of the linearization of the product of the price and the choice variables, and constraints (3.28) represent the valid inequalities.

$$
\begin{aligned}
Z^u = \max \quad & \frac{1}{R} \sum_{i \in \mathcal{C}_n \setminus \{0\}} \sum_{n=1}^{N} \sum_{r=1}^{R} \eta_{inr} && && (3.19) \\
\text{s.t.} \quad & U_{inr} = d_{inr} + \beta_{in} p_{in} && \forall i \in \mathcal{C}_n, n, r && (3.20) \\
& U_{inr} \leq U_{nr} && \forall i \in \mathcal{C}_n, n, r && (3.21) \\
& U_{nr} \leq U_{inr} + M_{inr}(1 - w_{inr}) && \forall i \in \mathcal{C}_n, n, r && (3.22) \\
& \sum_{i \in \mathcal{C}_n} w_{inr} = 1 && \forall n, r && (3.23) \\
& a_{in} w_{inr} \leq \eta_{inr} && \forall i \in \mathcal{C}_n \setminus \{0\}, n, r && (3.24) \\
& \eta_{inr} \leq b_{in} w_{inr} && \forall i \in \mathcal{C}_n \setminus \{0\}, n, r && (3.25) \\
& p_{in} - (1 - w_{inr}) b_{in} \leq \eta_{inr} && \forall i \in \mathcal{C}_n \setminus \{0\}, n, r && (3.26) \\
& \eta_{inr} \leq p_{in} - (1 - w_{inr}) a_{in} && \forall i \in \mathcal{C}_n \setminus \{0\}, n, r && (3.27) \\
& U_{inr} \leq d_{0nr} w_{0nr} + \sum_{j \in \mathcal{C}_n \setminus \{0\}} (\beta_{jn} \eta_{jnr} + d_{jnr} w_{jnr}) && \forall i \in \mathcal{C}_n, n, r && (3.28) \\
& w_{inr} \in \{0, 1\} && \forall i \in \mathcal{C}_n, n, r && (3.29)
\end{aligned}
$$

Model 3.2: Uncapacitated version of the revenue maximization problem (Model 3.1)

Finally, we notice that the general representation of the price through the variables p_{in} enables the definition of a different price for each individual, even if it might be quite unrealistic in practice. If this is the case, we notice that the complexity of the solution approach of the revenue maximization problem (Model 3.1) is notably reduced, as it is possible to iterate over the customers in the order determined by the priority list. For each customer, we solve the uncapacitated version of the problem, while keeping track of the available services to each individual. In other words, the capacity of each service is fully available at the beginning of each scenario, and it gets updated at every iteration, i.e., for every customer, based on the performed choices. As soon as the capacity is reached in a scenario, the service becomes unavailable for the upcoming customers in the priority list and that scenario. The total revenue is then obtained by aggregating the individual contributions.

The uncapacitated revenue maximization problem associated with individual n is included in Model 3.3. Notice that we introduce the sets $\mathcal{C}_{nr}$, which represent the set of services considered by individual n that are available at draw r, i.e., there is still remaining capacity to accommodate individual n. Notice that we could also incorporate availability variables to capture the availability at the draw level based on the choices of the preceding customers. Algorithm 3.1 presents the pseudocode of this procedure.

Nevertheless, the operator will typically propose different prices to different groups of customers with similar characteristics (e.g., students, seniors), or merely a single price to everyone. The price specification p_{in} also allows to capture these situations, which are more interesting not only because they are more common in real life, but also because

$$
\begin{aligned}
Z_n^u = \max \quad & \frac{1}{R} \sum_{i \in \mathcal{C}_{nr} \setminus \{0\}} \sum_{r=1}^{R} \eta_{inr} && && (3.30)\\
\text{s.t.} \quad & U_{inr} = d_{inr} + \beta_{in} p_{in} && \forall i \in \mathcal{C}_{nr}, r && (3.31)\\
& U_{inr} \leq U_{nr} && \forall i \in \mathcal{C}_{nr}, r && (3.32)\\
& U_{nr} \leq U_{inr} + M_{inr}(1 - w_{inr}) && \forall i \in \mathcal{C}_{nr}, r && (3.33)\\
& \sum_{i \in \mathcal{C}_{nr}} w_{inr} = 1 && \forall r && (3.34)\\
& a_{in} w_{inr} \leq \eta_{inr} && \forall i \in \mathcal{C}_{nr} \setminus \{0\}, r && (3.35)\\
& \eta_{inr} \leq b_{in} w_{inr} && \forall i \in \mathcal{C}_{nr} \setminus \{0\}, r && (3.36)\\
& p_{in} - (1 - w_{inr}) b_{in} \leq \eta_{inr} && \forall i \in \mathcal{C}_{nr} \setminus \{0\}, r && (3.37)\\
& \eta_{inr} \leq p_{in} - (1 - w_{inr}) a_{in} && \forall i \in \mathcal{C}_{nr} \setminus \{0\}, r && (3.38)\\
& U_{inr} \leq d_{0nr} w_{0nr} + \sum_{j \in \mathcal{C}_{nr} \setminus \{0\}} (\beta_{jn} \eta_{jnr} + d_{jnr} w_{jnr}) && \forall i \in \mathcal{C}_{nr}, r && (3.39)\\
& w_{inr} \in \{0, 1\} && \forall i \in \mathcal{C}_{nr}, r && (3.40)
\end{aligned}
$$

Model 3.3: Uncapacitated version of the revenue maximization problem (Model 3.1) associated with individual n

Algorithm 3.1: Solution method of the revenue maximization problem (Model 3.1) when p_{in} represent prices that are proposed at the individual level

Input: Revenue maximization problem (Model 3.1);
Output: Optimal solution of the revenue maximization problem (Model 3.1);

Initialize remaining capacity for each draw r: $\bar{c}_{ir} = c_i, \forall i \in \mathcal{C} \setminus \{0\}, r$;
Initialize the objective function $Z = 0$;
for $n = 1 \ldots N$ **do**
  for $r = 1 \ldots R$ **do**
    Define $\mathcal{C}_{nr} = \{0\}$;
    for $i \in \mathcal{C} \setminus \{0\}$ **do**
      if $\bar{c}_{ir} > 0$ *and* $i \in \mathcal{C}_n$ **then**
        $\mathcal{C}_{nr} = \mathcal{C}_{nr} \cup \{i\}$
  Solve Model 3.3 considering $\mathcal{C}_{nr}$ and obtain the values of the choice variables w_{inr} and the optimal objective function Z_n^u;
  $Z = Z + Z_n^u$;
  for $r = 1 \ldots R$ **do**
    for $i \in \mathcal{C} \setminus \{0\}$ **do**
      if $w_{inr} = 1$ **then**
        $\bar{c}_{ir} = c_i - 1$;

the revenue maximization problem cannot be solved in the way described above. In the experiments performed in Section 3.6 we assume that a single price is proposed to everyone, which hinders the use of procedures as the one described in Algorithm 3.1 to solve Model 3.1. In the following, we keep the general notation p_{in}, as the decomposition techniques here explored apply to all cases, but we are concerned with the price definition at the aggregated level.

3.3 Literature review

In order to favorably apply a decomposition technique, the optimization problem under consideration should have the appropriate decomposable structure. Two such structures arise in practice: one characterized by complicating variables and one characterized by complicating constraints. In real-world systems, investment decisions are typically integer, while the subsequent operation decisions are continuous. Dealing with integer variables is much more complicated than dealing with continuous variables, and once integer decisions have been made, the resulting subproblems usually decompose by blocks, which might facilitate their solution. The formulations associated with applications with decentralized structures typically present a set of constraints involving variables from all structures that prevent each of the corresponding subproblems to be solved separately.

These complicating elements are not mutually exclusive, and the same optimization problem can be seen as a problem with complicating variables or as a problem with complicating constraints. In either case, complicating variables and constraints make the problem more difficult to solve because they prevent an easy solution or a solution by blocks. When the complicating variables are fixed to some values, or the complicating constraints are ignored, the resulting problem decomposes in several simpler problems or acquires a structure that is less challenging to solve than the original one.

Sections 3.3.1 and 3.3.2 provide an overview of classical decomposition methods and hybrid algorithms and extensions thereof that have been widely investigated in the literature in the context of combinatorial optimization problems with complicating variables and constraints, respectively. Section 3.3.3 evaluates these techniques and assess their potential implementation for the choice-based optimization framework.

3.3.1 Complicating variables

The classical approach to handle problems containing complicating variables was proposed by Benders (1962). It has become one of the most commonly used exact algorithms for such problems because it exploits its structure and decentralizes the overall computational burden. In the Benders decomposition scheme, the model is first projected onto the subspace defined by the set of complicating variables. The resulting formulation is dualized, and the feasibility requirements (feasibility cuts) and the projected costs (optimality cuts) are determined by the associated extreme points and rays, respectively. Since the alternative formulation that enumerates all the extreme points and rays is computationally expensive, a relaxation strategy to the feasibility and optimality cuts is applied. This yields a master problem and subproblem(s) that are iteratively solved to generate the violated cuts.

When applied to MILP, the Benders decomposition procedure fixes the integer variables to given feasible integer values in order to solve the resulting continuous linear programming (LP) subproblem(s), for which standard duality theory can be used to develop cuts. This allows to obtain an upper bound on the optimal value of the objective function (for a minimization problem). The MILP master problem, which includes only a subset of the feasibility and optimality cuts, is solved to determine improved values of the integer variables. This yields a lower bound of the optimal value of the objective function. The method stops as soon as both bounds are close enough.

Generally, the MILP master problem is solved to optimality with branch-and-bound, and the subproblem(s) with the simplex method. As the former usually lacks a special structure and continually grows in size, a direct implementation of the classical Benders decomposition framework might require excessive computational time and memory (Rahmaniani et al., 2017). A significant body of research is dedicated to enhance the convergence of the algorithm by reducing two aspects: (i) the number of iterations, which has to do with the improvement of the quality of the generated solutions and cuts, and (ii) the time needed for each iteration, which is related to the solution procedure considered to optimize the master problem and the subproblem(s). Furthermore, the way in which the initial master problem and subproblem(s) are defined also has notable consequences on the efficiency of the algorithm. For example, in partial Benders decomposition (Crainic et al., 2016), explicit information associated with the non-complicating variables is added to the master problem. In the non-standard decomposition strategy by Gendron et al. (2016), the projected variables are retained in the master problem while relaxing the integrality requirements.

The number of iterations is closely related to the strength of the optimality and feasibility cuts and the quality of the generated solutions. The classical cut-generation scheme associated with solving the regular subproblem obtained from the decomposition might be inefficient, particularly when the subproblems are degenerated or infeasible. Maximal non-dominated cut generation may be more efficient to generate optimality cuts as they eliminate the need to solve the auxiliary problem. At the same time, feasibility cuts are found based on a random selection of the extreme rays, being the strategy that generates combinatorial cuts for subproblems with big-M constraints the only one that has proven its worth in practice. The quality and computational time of the generated solutions can be ameliorated with alternative formulations, by improving the master problem formulation with valid inequalities that strengthen the relaxed master problem, or by using heuristics to independently generate solutions or improve those already found.

The structure of the master problem and subproblem(s) can be exploited in order to improve the computational time at the iteration level. The size of the problem can be controlled by removing unnecessary cuts, and the optimality requirement at each iteration can be relaxed by solving, for instance, the master problem with single search tree instead of branch-and-bound. Furthermore, when the subproblem has a special structure, specialized algorithms are a better option than the simplex method.

3.3.2 Complicating constraints

The complicating constraints of a combinatorial optimization problem can be transferred to the objective function in a Lagrangian fashion, i.e., with Lagrangian multipliers. In this way, the original problem is simplified because its optimization over the set defined by the remaining constraints is relatively easy. This Lagrangian relaxation approach started to gain popularity after being studied by Geoffrion (1974), especially in the context of integer programming (IP). The first attempt as we know it today is the research in Held and Karp (1971a) and Held and Karp (1971b) on the traveling salesman problem.

The Lagrangian multipliers (or dual variables) associated with each constraint placed in the objective function penalize their violation. For any given values of these multipliers, the resulting relaxed problem, known as the Lagrangian subproblem, provides a lower bound (upper bound) on the optimal value of the objective function of the original minimization (maximization) problem. It provides a stronger (or at least equal) bound than the one generated by the LP relaxation of the original problem (if feasible), which is determined by dropping the integrality requirements. The equality is achieved when the Lagrangian relaxation has the integrality property, i.e., its solution (for any admissible values of the multipliers) is not altered by removing the integrality conditions on its variables (Geoffrion, 1974). In any event, it can be considered instead of the LP bound within a branch-and-bound algorithm, as well as to derive good feasible solutions.

The tightest of the Lagrangian bounds (for an original minimization problem) is obtained by solving the problem of maximizing the Lagrangian subproblem over the set of admissible Lagrangian multipliers. This problem is known as the Lagrangian dual because it coincides with the formal Lagrangian dual of the original problem with respect to the relaxed constraints (Geoffrion, 1971). The Lagrangian dual has a number of important structural properties that facilitate a hill climbing algorithm to find its solution (Fisher, 1981a). Its objective function (as a function of the Lagrangian multipliers) is the lower envelope of a finite family of linear functions, which makes it continuous and concave. Although it is differentiable almost everywhere, it is not differentiable at any multiplier where the associated Lagrangian subproblem has multiple optima.

A Lagrangian solution is an optimal solution of the Lagrangian subproblem for any given values of the Lagrangian multipliers. If a Lagrangian solution satisfies complementary slackness, it is an optimal solution of the IP problem (Guignard, 2003). If it is feasible but complementary slackness does not hold, it is a feasible solution of the original problem and it is still necessary to determine whether it is optimal or not. Notice that if the constraints being relaxed are equality constraints, complementary slackness holds automatically, and therefore feasibility of the Lagrangian solution for the original problem implies optimality.

Several extensions of Lagrangian relaxation can be found in the literature (Guignard, 2003), being Lagrangian decomposition (Guignard and Kim, 1987), or variable splitting, among the most popular ones. This technique artificially induces decomposition by introducing copies of the original variables for a subset of constraints of the original problem. It then dualizes the constraints that impose that these variables should be identical to the original ones. This approach is of special interest for problems whose constraint set is the intersection of several specially structured constraint sets. The Lagrangian decomposition bound can strictly dominate the Lagrangian relaxation bounds obtained by dualizing any set of constraints. Moreover, since the dualized constraints are equality constraints, if the Lagrangian subproblems (each on the corresponding set of copied variables) have the same optimal solution, then that solution is also optimal for the original MILP. More generally, Lagrangian substitution (Reinoso and Maculan, 1992) induces decomposition by creating more sophisticated substitutions than the copy constraint that imposes that the variables should be identical. This approach reduces the number of multipliers with respect to Lagrangian decomposition. Indeed, the number of constraints to be dualized decreases because some of them involve several variables. It generally yields bounds that are in between the corresponding Lagrangian relaxation and Lagrangian decomposition bounds.

Other strategies include augmented Lagrangian methods (Bertsekas, 1996), which have been mostly used in stochastic optimization and non-linear continuous programming, though they can also be used in linear and non-linear IP. The difference with respect to Lagrangian relaxation is the addition of a quadratic penalty term to the Lagrangian objective function, called augmentation. The major drawback of this approach is the loss of separability, for which several algorithmic approaches have been proposed (e.g., the Auxiliary Problem Principle by Cohen and Zhu, 1984).

Additionally, extensive research has been carried in order to come up with ways of modifying Lagrangian solutions to make them feasible. Lagrangian heuristics are essentially problem-dependent. They can attempt at modifying the solution to correct infeasibilities or be embedded at every iteration of the Lagrangian relaxation scheme to derive a feasible solution. However, if an optimal or almost optimal solution is desired, then branch-and-bound can be adapted by replacing the LP bounds by the Lagrangian ones.

Solving the Lagrangian dual is an important part of the Lagrangian relaxation method. This is why we find multiple approaches in the literature that exploit the structural properties of this problem to find optimal or near-optimal solutions (Guignard, 2003).

The subgradient method is an adaptation of the gradient method in which gradients are replaced by subgradients. As it is easy to implement and has worked well in many practical problems, it has become the most popular method to solve the Lagrangian dual, whose objective function is subdifferentiable everywhere. Its computational performance and theoretical convergence are assessed in Held et al. (1974), who show that the method converges to the optimal value of the Lagrangian dual if the sequence of step multipliers tends to 0 and the associated series tends to infinity. Unless we obtain a set of Lagrangian multipliers such that the optimal objective function of the Lagrangian dual is equal to that of a known feasible solution, there is no way of proving optimality, and the method should terminate upon meeting a certain stopping criterion (e.g., number of iterations). Some variants thereof have also been developed, such as the volume algorithm (Barahona and Anbil, 2000), which uses the information on all the previously generated Lagrangian solutions to determine the direction of motion.

Another class of algorithms to handle the Lagrangian dual consists of applying a variant of the simplex method that uses column generation techniques to the original problem. The goal is to generate an appropriate entering variable on each iteration by solving the Lagrangian subproblem with Lagrangian multipliers equal to the simplex multipliers (Fisher et al., 1975, Marsten et al., 1975). This approach is known to converge very slowly, and compared to the subgradient method, is harder to program and has not performed quite so well computationally (unless stabilization features are added). Multiplier adjustment methods (Erlenkotter, 1978, Fisher et al., 1986) are algorithms that specify the direction of motion with a finite and usually small set of primitive directions. For the implementation to be successful, such a set needs to be cleverly specified and should have a manageable size while still including enough ascent directions, which makes this approach problem specific.

Other strategies include the bundle method (Lemaréchal, 1989), cutting-plane methods (Kelley, 1960) and interior-point based solution approaches. The bundle method is a dual-ascent approach that solves a master problem at each iteration. The dual solution is often constrained to a given interval, and any deviation from the interval is penalized by a penalty function. As opposed to the subgradient method, it does not present a (potential) zigzagging behavior because in the worst case it provides a *null* step. The stabilized column generation approach can be understood as a bundle method applied on the dual of the restricted master problem with a polyhedral penalty function. Cutting-plane methods represent a standard non-linear programming approach to maximize a concave non-differentiable function which suffers from several drawbacks. Interior-point based solution approaches, such as the Analytic Center Cutting Plane Method (ACCPM), have also theoretically shown to have a better rate of convergence (Goffin and Vial, 2002). Combinations of the discussed procedures have also been implemented, such as a subgradient method in the first phase and a cutting-plane method in the second phase, or a combination of column generation with the subgradient method (Guignard, 2003).

A combinatorial problem with complicating constraints can be alternatively viewed as the problem of selecting a solution on the set defined by the simpler constraints that also satisfy the complicating constraints. This results in the so-called Dantzig-Wolfe reformulation (Dantzig and Wolfe, 1960). It replaces the original variables with a convex combination of the extreme points of the polyhedron corresponding to a substructure of the formulation. As the Lagrangian dual problem is the dual of the master problem associated with the Dantzig-Wolfe reformulation of an IP problem, both approaches provide the same bounds for the same original problem (Geoffrion, 1974). The corresponding equivalence is shown for Lagrangian decomposition (Guignard and Kim, 1987) when the Dantzig-Wolfe master problem is formulated on the variable duplicates.

The Dantzig-Wolfe reformulation of an IP problem gives rise to an IP master problem that is solved to optimality with the algorithm known as branch-and-price, which combines column generation with branch-and-bound. There are many practical issues arising when developing a branch-and-price algorithm, such as stabilization of the column generation procedure (of the associated LP problems) or branching strategies. As discussed in Vanderbeck (2000), there is a trade-off between branching efficiency and subproblem tractability.

3.3.3 Discussion

Sections 3.3.1 and 3.3.2 clearly show that the literature on decomposition techniques is immense. As the same formulation can be manipulated from different angles, multiple strategies can be tailored to the problem of interest. Despite this lack of generalization, some comprehensive guidance is available and can be taken into consideration while building a decomposition scheme.

In the state-of-the-art survey presented in Rahmaniani et al. (2017), the Benders decomposition method appears to be particularly appropriate for problems with few complicating variables (usually binary) and so many continuous variables that solving the problem as a whole is inefficient. There are many examples of such problems in stochastic programming. Benders decomposition methods adapted to such problems (originally proposed by Van Slyke and Wets, 1969 for stochastic LP problems) are generally known as primal decomposition methods. They are usually illustrated in the two-stage stochastic programming problem, where the set of decision variables is composed by first-stage decisions, which are deterministic, and the second-stage (recourse) decisions, which are allowed to depend on the random problem data (Ruszczyński, 2003). In the case of mixed-integer two-stage problems (Wollmer, 1980, Laporte and Louveaux, 1993), the part of the objective function on the variables that have not been fixed (the expected value function in the two-stage model) is non-convex, so that it cannot be described using linear cuts as in the continuous case. However, when the first-stage variables are binary, it is possible to define a valid set of linear optimality cuts, which are iteratively generated in a branch-and-cut scheme as there can be a large number of such cuts. When a finite number of scenarios (realizations of the random data in the two-stage setting) is assumed, and therefore the deterministic equivalent formulation that relies on all scenarios and the associated probabilities can be considered, these methods generally lead to master problems that are governed by non-convex and non-differentiable functions of the same type as the value function of an IP (Carøe and Tind, 1998).

Additionally, the Benders decomposition method can handle problems with weak linear relaxations and numerical instability as a result of big-M constraints and the associated binary variables by transferring such constraints to the subproblem(s) and using specialized cuts to represent them. More generally, the Benders decomposition method has been convenient for problems in which temporarily fixing the complicating variables makes the resulting problem significantly easier to handle by, for instance, transforming a non-convex problem into a convex one.

There are often many ways in which a given problem can be relaxed in a Lagrangian fashion (Guignard, 2003). The most commonly used approach consists of isolating an interesting subproblem by dualizing the non-related constraints. The Lagrangian subproblems are interesting because they have a special structure that can be exploited and there might even be specialized algorithms for solving them efficiently. Lagrangian decomposition can be applied if there is more than one interesting subproblem with common variables. Instead of relaxing the individual copies of variables, it is also possible to apply Lagrangian substitution and dualize the aggregate copy constraints associated with the variable duplicates. Another usual strategy involves the relaxation of the linking constraints, i.e., the constraints that couple together independent structures that are contained in the same problem. Regardless of the variant, the Lagrangian subproblem commonly decomposes into smaller problems, which prompts a decrease in the computational complexity, as it is generally easier to solve a larger number of problems with a reduced number of binary variables than a single problem with many binary variables.

Lagrangian relaxation strategies are as well considered in stochastic programming, and are known as dual decomposition methods. Whereas primal decomposition methods decompose the problem by time stages and operate by searching for increasingly better solutions, dual decomposition methods consider subproblems associated with scenarios and are governed by finding good dual multipliers in an iterative basis by solving the Lagrangian dual (Haneveld and van der Vlerk, 1999). Scenario decomposition for large multistage stochastic programming problems in the continuous case was proposed by Mulvey and Ruszczyński (1995). Carøe and Schultz (1999) characterize a scenario decomposition algorithm for two-stage problems with mixed-integer variables in both stages. The authors assume a finite number of scenarios and consider the deterministic equivalent formulation of the problem. The principle is that of Lagrangian decomposition. Indeed, copies of the first-stage variables, which do not depend on the random data, are introduced for each scenario and the constraints that impose that such variables should not depend on the scenario, known as non-anticipativity constraints, are relaxed in a Lagrangian manner. A branch-and-bound that uses Lagrangian relaxation of the non-anticipativity constraints as a bounding procedure is proposed to achieve convergence. Dual decomposition methods differ with respect to the formulation of non-anticipativity constraints, the definition of the Lagrangian, the construction of the subproblems and the way the multipliers are updated. Furthermore, non-anticipativity constraints can be seen as hard constraints because they couple constraints associated with different scenarios. Motivated by the high dimension of the vector of Lagrangian multipliers, and based on duality results involving augmented Lagrangian (see Section 3.3.2), algorithms like progressive hedging (e.g., Rockafellar and Wets, 1991 for multi-stage mixed-integer 0-1 problems) and the Jacobi method (e.g., Rosa and Ruszczyński, 1996 for multi-stage convex problems) have been developed and applied to a variety of problems.

Concerning the solving of the Lagrangian dual, Frangioni et al. (2017) conclude that the subgradient method can be competitive with more sophisticated approaches to handle the Lagrangian dual when the tolerance required for the solution is not too tight. This is the case when solving the Lagrangian dual of MILP. The requirement is to have appropriate tuned parameters, which they describe as a not extremely difficult task.

The choice-based optimization framework, and particularly the revenue maximization problem (Model 3.1), has two sets of binary variables: the choice variables, which are the result of the linearization of the variable capturing the highest discounted utility for a given customer and scenario, and the availability variables at the scenario level, which are involved in the linearization of the discounted utility and the capacity constraints. Furthermore, both variables are related in constraints (3.10), as a service that is not available cannot be chosen.

Various complicating constraints can be identified. The utility expression (3.2) works as the link between the DCM (customers) and the MILP model (operator). The capacity constraints (3.11)–(3.12) complicate the solution approach and prevent a potential decomposition by customer. The model also relies on big-M constraints to linearize the definition of the discounted utility (constraints (3.3)–(3.6)) and the maximization of the highest discounted utility (constraints (3.7)–(3.9)).

Since the big-M constraints are the tightest possible to our formulation, and the constraints mentioned above do not allow to exploit the two decomposition dimensions discerned in Section 3.1 (the customers and the scenarios), we resolve to rely on Lagrangian relaxation to take advantage of the decomposable structure of the choice-based optimization framework. To this end, Section 3.4 reviews the initial attempts to decompose the revenue maximization problem, and Section 3.5 presents the proposed Lagrangian decomposition scheme, which is inspired by the scenario decomposition method for stochastic programing. The developed approach is built on a subgradient method that integrates the solution to the Lagrangian subproblems and a heuristic to derive feasible solutions to the revenue maximization problem.

3.4 Initial attempts on the Lagrangian relaxation scheme

In this section, we summarize the first approaches based on Lagrangian relaxation that were developed to induce separability in the choice-based optimization framework. In Section 3.4.1, we describe the decomposition strategy that allows to isolate two interesting subproblems that can be identified concerning the optimizer agent: one related to the customers and one related to the operator. The former comprises the selection of the service maximizing their utility, and the latter consists of the definition of a pricing strategy that maximizes the expected revenue. This method yields trivial solutions of the resulting subproblems, and hinders the derivation of feasible solutions to the original problem.

With the aim of preserving the constraints linking the formulation blocks associated with the operator and the customers, and motivated by the main idea in scenario decomposition in stochastic programming, we rely exclusively on Lagrangian decomposition on the price variables (the only set of variables that do not depend on the scenarios) in Section 3.4.2. Such decomposition is later generalized in Section 3.5, as the splitting of price variables for subsets of scenarios (rather than single scenarios) addresses one of the identified limitations with respect to the obtained Lagrangian solution.

3.4.1 The customer and the operator subproblems

One of the common strategies pointed out in Section 3.3.3 consists of the relaxation of the constraints that tie independent structures within the problem. Two such structures can be distinguished in the revenue maximization problem, each of them related to an optimizer agent: the customer and the operator. Both structures are linked via the utility constraints (3.20). Indeed, the utility specification associates a score with each service, which characterizes the behavior of the customers as assumed by the operator (governed by a DCM), and determines the choice for each individual and simulation draw. Furthermore, the utility captures the sensitivity of customers towards price, which is considered by the operator to decide on the prices to propose. We describe the resulting decomposition strategy with the uncapacitated version of the revenue maximization problem (Model 3.2).

Relaxation

The mixed-integer linear formulation for the DCM, i.e., constraints (3.20)–(3.23) and (3.29), determine the problem associated with the customers. Provided that the prices are known, such constraints assign values to the utility variables and define the choices associated with each customer and draw. The problem associated with the operator consists of the revenue maximization problem that does not include the disaggregate demand representation, i.e., constraints (3.24)–(3.27), and the objective function (3.19). Therefore, the customer subproblem involves the variables U_{inr}, U_{nr}, p_{in} and w_{inr}, and the operator subproblem involves p_{in}, w_{inr} and η_{inr}.

In order to separate both subproblems, it is necessary to designate an independent set of variables to each of them. Note that the valid inequalities (3.28) relate variables of both subproblems, which prevents the decomposition. Since such constraints are equivalent to (3.21)–(3.22), and their impact on computational savings for the Lagrangian subproblem is expected to be less relevant than for the original problem, we decide to ignore them in this decomposition strategy.

The decomposition in a Lagrangian fashion is constructed in two steps. First, the utility constraints (3.20) are transferred to the objective function with associated Lagrangian multipliers $\rho_{inr} \in \mathbb{R}, \forall i \in \mathcal{C}_n, n, r$. Notice that the penalty factors are unrestricted in sign because the relaxed constraints are equality constraints. Second, we introduce duplicates v_{inr} of the choice variables $w_{inr}, \forall i \in \mathcal{C}_n, n, r$, and dualize the constraints that relate both sets of variables. Instead of writing the constraints that impose that these duplicates should be identical to the original variables as equality constraints, we take advantage of the structure of the problem and represent them with the following equivalent constraints:

$$v_{inr} \leq w_{inr}, \qquad \forall i \in \mathcal{C}_n, n, r, \tag{3.41}$$

$$\sum_{i \in \mathcal{C}_n} v_{inr} = 1, \qquad \forall n, r, \tag{3.42}$$

$$v_{inr} \in \{0, 1\}, \qquad \forall i \in \mathcal{C}_n, n, r. \tag{3.43}$$

The advantage of this formulation is the insertion of the redundant assignment constraints (3.42), which strengthens the Lagrangian subproblem. Constraints (3.41) are then relaxed with Lagrangian multipliers $\psi_{inr} \in \mathbb{R}_{\geq 0}, \forall i \in \mathcal{C}_n, n, r$.

Additionally, when the utility variables are dissociated from the price variables, the bounds on the utility derived from the price bounds need to be stated explicitly:

$$\ell_{inr} \leq U_{inr} \leq m_{inr}, \qquad \forall i \in \mathcal{C}_n, n, r. \tag{3.44}$$

The Lagrangian subproblem obtained from the relaxation of constraints (3.20) and (3.41) is presented in Model 3.4. This subproblem further splits into the operator and customer subproblems. Note that the term $\sum_{i \in \mathcal{C}_n} \sum_{n=1}^{N} \sum_{r=1}^{R} -\rho_{inr} d_{inr}$ is not included in the objective function of any of the subproblems as it is constant to the optimization.

$$
\begin{aligned}
Z^{LR}(\rho,\psi) = \max \quad & \sum_{i\in\mathcal{C}_n\setminus\{0\}} \sum_{n=1}^{N}\sum_{r=1}^{R} \frac{1}{R}\eta_{inr} \\
& + \sum_{i\in\mathcal{C}_n} \sum_{n=1}^{N}\sum_{r=1}^{R} \rho_{inr}(U_{inr} - d_{inr} - \beta_{in}p_{in}) \\
& + \sum_{i\in\mathcal{C}_n} \sum_{n=1}^{N}\sum_{r=1}^{R} \psi_{inr}(v_{inr} - w_{inr}) & (3.45) \\
\text{s.t.} \quad & (3.21) - (3.23), (3.29), (3.44) \text{ [customer subproblem]} \\
& (3.24) - (3.27), (3.42) - (3.43) \text{ [operator subproblem]}
\end{aligned}
$$

Model 3.4: Lagrangian subproblem for the Lagrangian relaxation on the utility constraints for the uncapacitated version of the revenue maximization problem (Model 3.2)

Customer subproblem

The customer subproblem decomposes by individual and scenario. It is detailed in Model 3.5. Notice that the value of U_{inr} can only be ℓ_{inr} (if $\rho_{inr} \leq 0$) or m_{inr} (if $\rho_{inr} \geq 0$) in order to ensure the highest possible contribution to the objective function. The choice variable w_{inr} associated with the service with the highest value for U_{inr} is then set to 1, and if several services share such value, the one with the highest contribution to the objective function, i.e., $\rho_{inr}U_{inr} - \psi_{inr}w_{inr}$, is selected.

$$
\begin{aligned}
Z^{LR,u}_{nr}(\rho,\psi) = \max \quad & \sum_{i\in\mathcal{C}_n} (\rho_{inr}U_{inr} - \psi_{inr}w_{inr}) & & (3.46) \\
\text{s.t.} \quad & U_{inr} \leq U_{nr} & \forall i \in \mathcal{C}_n & \quad (3.21) \\
& U_{nr} \leq U_{inr} + M_{inr}(1 - w_{inr}) & \forall i \in \mathcal{C}_n & \quad (3.22) \\
& \sum_{i\in\mathcal{C}_n} w_{inr} = 1 & & \quad (3.23) \\
& w_{inr} \in \{0,1\} & \forall i \in \mathcal{C}_n & \quad (3.29) \\
& \ell_{inr} \leq U_{inr} \leq m_{inr} & \forall i \in \mathcal{C}_n & \quad (3.44)
\end{aligned}
$$

Model 3.5: Customer subproblem associated with individual n and scenario r for the Lagrangian subproblem in Model 3.4

Operator subproblem

Model 3.6 formulates the operator subproblem, which only decomposes by individual. A decomposition by scenario is not possible because the price needs to be the same across draws. It is easy to see that for each configuration of the choice variables $\{v_{inr}\}_{inr}$ satisfying $\sum_{i\in\mathcal{C}_n} v_{inr} = 1, \forall n, r$, the objective function associated with individual n is an affine function on p_{in}, which implies that the price variables can only take values equal to the price bounds, i.e., p_{in} is either equal to a_{in} or b_{in}.

$$
\begin{aligned}
Z_n^{LR,o}(\rho,\psi) = \max \quad & \sum_{i\in\mathcal{C}_n\setminus\{0\}} \sum_{r=1}^{R} \left(\frac{1}{R}\eta_{inr} - \rho_{inr}\beta_{in}p_{in}\right) && \\
& + \sum_{i\in\mathcal{C}_n} \sum_{r=1}^{R} \psi_{inr} v_{inr} && (3.47) \\
\text{s.t.} \quad & a_{in}v_{inr} \le \eta_{inr} & \forall i \in \mathcal{C}_n \setminus \{0\}, r \quad & (3.24) \\
& \eta_{inr} \le b_{in}v_{inr} & \forall i \in \mathcal{C}_n \setminus \{0\}, r \quad & (3.25) \\
& p_{in} - (1 - v_{inr})b_{in} \le \eta_{inr} & \forall i \in \mathcal{C}_n \setminus \{0\}, r \quad & (3.26) \\
& \eta_{inr} \le p_{in} - (1 - v_{inr})a_{in} & \forall i \in \mathcal{C}_n \setminus \{0\}, r \quad & (3.27) \\
& \sum_{i\in\mathcal{C}_n} v_{inr} = 1 & \forall r \quad & (3.42) \\
& v_{inr} \in \{0,1\} & \forall i \in \mathcal{C}_n, r \quad & (3.43)
\end{aligned}
$$

Model 3.6: Operator subproblem associated with individual n for the Lagrangian subproblem in Model 3.4

Discussion

Even though the resulting subproblems are less complex to solve than the original one, the upper bound provided by the Lagrangian subproblem is restricted, in the sense that both p_{in} and U_{inr} can only take the corresponding extreme values. In practice, however, this is not the case, and in-between values of both variables might be achieved. Hence, the irrelevant nature of the Lagrangian solutions yields to trivial feasible solutions of the original problem.

Consequently, the utility constraints (3.20), which establish the link between the customers and the operator, should not be dualized, and need to be preserved in the decomposed structures. The implementation of both subproblems for given values of the multipliers showed the behavior previously described. This limits the integration of this decomposition strategy in an iterative method such as the subgradient method (see Section 3.5.3), as the directions of motion derived from the obtained solutions become trivial.

We resolve to rely exclusively on Lagrangian decomposition in Section 3.4.2. The splitting of the choice variables, which are the only discrete variables in the formulation, is meaningless if the utility constraints are not dualized because neither independent sets of variables can be identified nor the resulting Lagrangian subproblem is provided with any decomposable structure to be exploited. Instead, we induce separability by splitting the price variables across draws, as they are the only variables in Model 3.2 that prevent its decomposition at the scenario level.

3.4.2 Lagrangian decomposition on the price variables

The uncapacitated version of the revenue maximization problem (Model 3.2) breaks up into N problems without the need to relax the formulation. Notice that if the prices are proposed at the group level, the problem splits into as many problems as groups of individuals are considered, i.e., as many as different prices are being proposed. The decomposition by scenarios, however, is not possible because the price variables p_{in} need to be the same across draws. The remaining variables (U_{inr}, U_{nr}, w_{inr} and η_{inr}) and all constraints in Model 3.2 only involve a single scenario at a time, which means that if the price was also defined for each scenario, the subproblem associated with customer n will in turn decompose into R subproblems.

Relaxation

In order to prompt such decomposition, we create R copies of the variables p_{in} (one for each draw). We then dualize the constraints that impose that these copies must be equal to each other (the equivalent to non-anticipativity constraints in dual decomposition methods in stochastic programming), such that they are equivalent to the original variables. In this way, the original structure of the problem is preserved in the Lagrangian subproblem because all the original constraints are kept. We denote by $p_{inr} \in \mathbb{R}, \forall i \in \mathcal{C}_n \setminus \{0\}, n, r$ the copy of p_{in} associated with scenario r. These duplicates are also bounded, i.e., $a_{in} \leq p_{inr} \leq b_{in}, \forall i \in \mathcal{C}_n \setminus \{0\}, n, r$, and they are related to each other by means of the following equality constraints:

$$p_{inr} - p_{in(r+1)} = 0, \qquad \forall i \in \mathcal{C}_n \setminus \{0\}, n, r < R, \tag{3.48}$$

$$p_{inr} - p_{in1} = 0, \qquad \forall i \in \mathcal{C}_n \setminus \{0\}, n, r = R. \tag{3.49}$$

Note that constraints (3.49) are redundant, as the sequence of constraints (3.48) already ensures $p_{in1} = p_{inR}$. For the sake of completeness, we have decided to include them in order to obtain as many constraints as price copies.

Solving Model 3.2 is equivalent to solve its reformulation in terms of p_{inr}, which consists of replacing p_{in} by p_{inr} in constraints (3.20) and (3.26)–(3.27) and adding constraints (3.48)–(3.49) to ensure that the price is the same across draws.

We dualize constraints (3.48)–(3.49) with associated multipliers $\alpha_{inr} \in \mathbb{R}, \forall i \in \mathcal{C}_n \setminus \{0\}, n, r$. The resulting Lagrangian subproblem decomposes by customer and scenario.

Subproblems

The subproblem associated with customer n and scenario r is included in Model 3.7. Notice that α_{in0} (obtained when $r = 1$) refers to α_{inR}. The objective function (3.50) arises from a simple mathematical manipulation of the relaxed constraints, as each price copy

p_{inr} always appears in two constraints: once with positive sign and associated multiplier α_{inr}, and once with negative sign and associated multiplier $\alpha_{in(r-1)}$. As opposed to the Lagrangian relaxation scheme constructed in Section 3.4.1, the valid inequality (3.60) can be kept in the formulation.

$$
\begin{aligned}
Z^{LD}_{nr}(\alpha) = \max \quad & \sum_{i \in \mathcal{C}_n \setminus \{0\}} \left[\frac{1}{R} \eta_{inr} + (\alpha_{inr} - \alpha_{in(r-1)}) p_{inr} \right] && && (3.50) \\
\text{s.t.} \quad & U_{inr} = \beta_{in} p_{inr} + d_{inr} && \forall i \in \mathcal{C}_n && (3.51) \\
& U_{inr} \leq U_{nr} && \forall i \in \mathcal{C}_n && (3.52) \\
& U_{nr} \leq U_{inr} + M_{nr}(1 - w_{inr}) && \forall i \in \mathcal{C}_n && (3.53) \\
& \sum_{i \in \mathcal{C}} w_{inr} = 1 && && (3.54) \\
& a_{in} w_{inr} \leq \eta_{inr} && \forall i \in \mathcal{C}_n \setminus \{0\} && (3.55) \\
& \eta_{inr} \leq b_{in} w_{inr} && \forall i \in \mathcal{C}_n \setminus \{0\} && (3.56) \\
& p_{inr} - (1 - w_{inr}) b_{in} \leq \eta_{inr} && \forall i \in \mathcal{C}_n \setminus \{0\} && (3.57) \\
& \eta_{inr} \leq p_{inr} - (1 - w_{inr}) a_{in} && \forall i \in \mathcal{C}_n \setminus \{0\} && (3.58) \\
& U_{inr} \leq d_{0nr} w_{0nr} + \sum_{j \in \mathcal{C}_n \setminus \{0\}} (\beta_{jn} \eta_{jnr} + d_{jnr} w_{jnr}) && \forall i \in \mathcal{C}_n && (3.59) \\
& w_{inr} \in \{0, 1\} && \forall i \in \mathcal{C}_n && (3.60)
\end{aligned}
$$

Model 3.7: Lagrangian subproblem associated with individual n and scenario r for the Lagrangian decomposition on the price variables for the uncapacitated version of the revenue maximization problem (Model 3.2)

Such a subproblem needs to be solved for all individuals and scenarios under consideration. We denote by $Z^{LD}(\alpha)$ the Lagrangian subproblem that aggregates the subproblems for all individuals and scenarios:

$$
Z^{LD}(\alpha) = \sum_{n=1}^{N} \sum_{r=1}^{R} Z^{LD}_{nr}(\alpha). \tag{3.61}
$$

For given values of α_{inr}, Model 3.7 is a MILP formulation whose only integer variables are the binary variables w_{inr}. As the number of services is typically small, we address the combinatorial nature of this problem with enumeration. More precisely, for each individual n and draw r we iterate over the services in $\mathcal{C}_n$. At each iteration we assume that service $j \in \mathcal{C}_n$ is chosen, i.e., $w_{jnr} = 1$ and $w_{inr} = 0, \forall i \in \mathcal{C}_n \setminus \{j\}$, and solve the problem for the remaining variables. The service providing the highest contribution to the objective function is then set as the choice for that customer at that draw.

Under this assumption, we can dispense with the binary variables w_{inr}, and Model 3.7 becomes Model 3.8, which represents an LP formulation. As the choice variables have fixed values, the variables η_{inr} and the associated constraints (3.55)–(3.58) are not required, and can be replaced by constraints (3.65) to simply set the bounds on the price variables. Moreover, $U_{nr} = U_{jnr}$ (because $w_{jnr} = 1$), and therefore we can directly rely on the valid inequalities (3.59) to impose that U_{jnr} is the largest among all services. Note that constraints (3.54) are automatically satisfied by the fact that at each iteration we enforce that only one service is selected per customer and draw.

$$
\begin{aligned}
Z^{LD}_{jnr}(\alpha) = \max \quad & \delta_{jnr} p_{jnr} + \sum_{i \in \mathcal{C}_n \setminus \{0,j\}} (\alpha_{inr} - \alpha_{in(r-1)}) p_{inr} && && (3.62)\\
\text{s.t.} \quad & U_{inr} = \beta_{in} p_{inr} + d_{inr} && \forall i \in \mathcal{C}_n && (3.63)\\
& U_{jnr} \geq U_{inr} && \forall i \in \mathcal{C}_n \setminus \{j\} && (3.64)\\
& a_{in} \leq p_{inr} \leq b_{in} && \forall i \in \mathcal{C}_n \setminus \{0\} && (3.65)\\
\text{where} \quad & \delta_{jnr} = \begin{cases} 0 & \text{if } j = 0 \\ \frac{1}{R} + \alpha_{jnr} - \alpha_{jn(r-1)} & \text{otherwise} \end{cases}
\end{aligned}
$$

Model 3.8: LP problem for the Lagrangian subproblem (Model 3.7) when service j is assumed to be chosen

Notice that the problem formulated in Model 3.8 might be infeasible because the price p_{jnr} guaranteeing that U_{jnr} is the largest among the other services in $\mathcal{C}_n$ might lie outside the range set by constraints (3.65). However, by construction, there is at least one service $i \in \mathcal{C}_n$ for which the problem formulated by Model 3.7 is feasible. Indeed, for any given values of the variables $p_{inr}, \forall i \in \mathcal{C}_n \setminus \{0\}$, satisfying constraints (3.65), we can calculate the associated utility values U_{inr} and select the service $j \in \mathcal{C}_n$ achieving the highest utility, i.e., $w_{jnr} = 1$ and $w_{inr} = 0, \forall i \in \mathcal{C}_n \setminus \{j\}$. Expressed differently, it is always possible to find a feasible price configuration for at least one of the services assumed to be chosen by customer n at scenario r.

Discussion

Let us assume that the sensitivity towards price is negative (i.e., $\beta_{in} \leq 0, \forall i \in \mathcal{C}_n \setminus \{0\}, n$). This is typically the case, as the larger the price, the lower the attraction of the customer to the service. Given a feasible problem formulated by Model 3.8, the price associated with unchosen service $i \in \mathcal{C}_n \setminus \{0\}$ can be characterized as follows:

$$
p_{inr} = \begin{cases} \max\{a_{in}, p^*_{inr}\}, & \text{if } \alpha_{inr} - \alpha_{in(r-1)} \leq 0, \\ b_{in}, & \text{otherwise,} \end{cases} \quad \forall i \in \mathcal{C}_n \setminus \{0\} \mid w_{inr} = 0, \quad (3.66)
$$

where $p^*_{inr} \in \mathbb{R}$ is such that $U_{inr} = U_{jnr}$. As the subproblem is feasible, there exists a price p_{jnr} satisfying constraints (3.65) such that $U_{jnr} \geq U_{inr}, \forall i \in \mathcal{C}_n \setminus \{j\}$. The price p^*_{inr} can be graphically pictured at the intersection between U_{inr} and U_{jnr} evaluated at the optimal price (see Figure 3.1). If such intersection takes place outside the price range of service i, then it is assigned to a_{in} or b_{in} based on the sign of the associated difference of multipliers. Note that when this difference is positive, p_{inr} can be simply set to b_{in} because it provides the largest possible contribution to the objective function while ensuring that service j has the largest utility (thanks to the assumption on β_{in}).

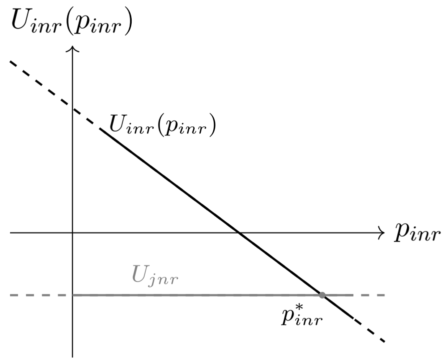

Figure 3.1: Graphical representation of p^*_{inr}

Thus, when the prices are defined at the individual level, the price associated with an unchosen service might be set to one of the bounds. This has an impact on the subgradient method, as the direction of motion to update the multipliers α_{inr} at each iteration is calculated in terms of the subgradient, which is defined as the constraints being dualized, i.e., $p_{inr} - p_{in(r+1)}$, evaluated at the values obtained after solving the corresponding subproblem (see Section 3.5.3 for further details). If any of these variables (or both) are equal to a price bound, the subgradient becomes meaningless, which hinders the computation of good bounds and slows down the convergence of the method.

Note that a price characterization for groups of customers may reduce such behavior because the choices of the individuals within the same group are simultaneously taken into account to determine the prices. Hence, there might be different services being chosen in the same subproblem, or even if a service is not chosen at all, its price needs to guarantee that the utilities corresponding to the chosen services are the highest, which might contribute to associate a relevant value with it.

After performing several tests on this decomposition scheme, we have observed that the characterization of prices induced by the Lagrangian subproblems leads to irrelevant directions of motion as dictated by the subgradient method (see Section 3.5.3). Different definitions of the direction of motion and various paradigms to update the step size were considered within the subgradient method. Nevertheless, this had limited impact on the quality of the obtained bounds due to the nature of the Lagrangian subproblems.

Furthermore, the definition of a copy for each draw generates a high number of variables that need to be reconciled throughout the subgradient method, which has a relevant impact on its convergence. In Section 3.5, we generalize this decomposition technique with the aim of getting over these difficulties.

3.5 Lagrangian decomposition scheme

The Lagrangian decomposition approach described in Section 3.4.2 gives rise to a relaxation that is too disaggregate, especially when a different price is proposed to each customer. Indeed, the splitting of the price variables p_{in} at the draw level yields a large number of constraints to be dualized, and in consequence, a large number of Lagrangian multipliers to be updated. This affects not only the convergence of the approach, but also the quality of the solution, as the price variables p_{inr} might take irrelevant values throughout the algorithm.

In order to address these limitations, we generalize the strategy by relying on copies of the price variables p_{in} associated with subsets of scenarios (instead of single scenarios). This decomposition strategy has lately received increased attention in stochastic programming after recent studies have shown that it allows to improve the obtained bounds while reducing the computational burden of solving the Lagrangian dual (Escudero et al., 2016, Crainic et al., 2014, Escudero et al., 2013). In this way, a lower number of constraints needs to be relaxed, which means that the decomposition framework is strengthened. Moreover, the more scenarios are simultaneously involved, the more likely it is to obtain more meaningful price copies, and it is therefore less likely to end up with copies that are equal to the price bounds.

We describe in detail the three components of the Lagrangian decomposition scheme in the following. Section 3.5.1 presents the Lagrangian decomposition involving multiple draws for the (capacitated) revenue maximization problem in Model 3.1, Section 3.5.2 explains the heuristic strategy to compute feasible solutions, and Section 3.5.3 integrates the decomposition and the heuristic within the subgradient method.

3.5.1 Decomposition by subsets of draws

In contrast to Section 3.4, we characterize the Lagrangian decomposition framework directly for the revenue maximization problem formulated by Model 3.1. Such problem does not split by customer due to the capacity constraints (3.11)–(3.12). However, as discussed in Section 3.2, if p_{in} is proposed to each individual, it is possible to ignore them and solve the uncapacitated version of the problem associated with each individual in an iterative manner while keeping track of the occupancy of the services. We are concerned with the more general case, where the price is defined for groups of customers with similar characteristics, or even a single price is proposed to everyone (as is done in Section 3.6), as it is more common in practice and more difficult to handle.

As pointed out in Section 3.4.2, we construct a Lagrangian decomposition scheme to induce separability with respect to the draw dimension by splitting the variables p_{in} among subsets of draws. We consider S subsets of draws, each of them indexed by s and denoted by R_s. With the aim of balancing the number of scenarios per subset, we define σ subsets with $\lceil R/S \rceil$ draws and $S - \sigma$ subsets with $\lfloor S/R \rfloor$ draws, where σ is the remainder in the Euclidean division of R by S. For each subset s, we denote by $p^s_{in} \in \mathbb{R}, a_{in} \leq p^s_{in} \leq b_{in}$, the associated copy of $p_{in}, \forall i \in \mathcal{C}_n \setminus \{0\}, n$. Constraints (3.67)–(3.68) impose equality among copies. We also include redundant constraints (3.68) for the sake of completeness.

$$p^s_{in} - p^{(s+1)}_{in} = 0, \qquad \forall i \in \mathcal{C}_n \setminus \{0\}, n, s < S, \tag{3.67}$$

$$p^s_{in} - p^1_{in} = 0, \qquad \forall i \in \mathcal{C}_n \setminus \{0\}, n, s = S. \tag{3.68}$$

The relaxation of constraints (3.67)–(3.68) in a Lagrangian fashion with associated multipliers $\alpha^s_{in} \in \mathbb{R}, \forall i \in \mathcal{C}_n \setminus \{0\}, n, s$, yields a Lagrangian subproblem that decomposes into independent subproblems for each subset s, as shown by Model 3.9. Notice that α^0_{in} (obtained when $s = 1$) refers to α^S_{in}. Once such subproblems are solved for all subsets s, we can compute the upper bound on Z as follows:

$$Z^{UB}(\alpha) = \sum_{s=1}^{S} Z^{UB}_s(\alpha), \tag{3.69}$$

where $Z^{UB}(\alpha)$ denotes the Lagrangian subproblem associated with the revenue maximization problem, which provides an upper bound on Z (Model 3.1) for any admissible values of the Lagrangian multipliers, i.e., $Z^{UB}(\alpha) \geq Z, \forall \alpha$.

Model 3.9 is essentially Model 3.1 for a reduced number of draws. As illustrated in Section 2.4, the computational complexity does not linearly grow along the draw dimension, which means that for a given number of individuals N, it might be more efficient to solve multiple problems with a smaller number of scenarios each than a single problem that contains all the scenarios. Moreover, the number of subsets S can be decided by the analyst based on the requirements with respect to computational time and quality of the solution, which provides a great deal of flexibility to the approach.

This methodology is a generalization of the decomposition strategy introduced in Section 3.4.2 that generates copies of the prices at the draw level because we can still define $S = R$ to have a single draw per subset. Nevertheless, the fact of simultaneously considering several draws might helps to address the difficulties previously identified. Indeed, if the price is individually defined, a problem involving multiple draws can help to increase the variety of choices within the problem, which might yield more meaningful prices. In the more general case, the grouping of the draws might also provide better Lagrangian solutions as each problem also contains more information.

Since the computational time grows exponentially with respect to the number of draws, it is important to assess the size of the subsets. If they contain a low number of draws, the resulting subproblems will be small in size, and therefore computationally less complex.

However, more iterations of the subgradient method might be needed to reach the same duality gap that is achieved in less iterations with subproblems with a higher number of draws each. Hence, there exists a trade-off between the size of the subsets, the computational complexity and the quality of the solution of the Lagrangian subproblems, and the number of iterations of the subgradient method. This is extensively analyzed in Section 3.6.2.

$$
\begin{aligned}
Z_s^{UB}(\alpha) = \max \quad & \sum_{i\in\mathcal{C}\setminus\{0\}} \sum_{n=1}^{N} \sum_{r\in R_s} \frac{1}{R}\eta_{inr} && (3.70)\\
& + \sum_{i\in\mathcal{C}\setminus\{0\}} \sum_{n=1}^{N} (\alpha_{in}^{s} - \alpha_{in}^{s-1}) p_{in}^{s} \\
\text{s.t.} \quad & U_{inr} = \beta_{in} p_{in}^{s} + d_{inr} && \forall i \in \mathcal{C}, n, r \in R_s \quad (3.71)\\
& \ell_{nr} \leq z_{inr} && \forall i \in \mathcal{C}, n, r \in R_s \quad (3.72)\\
& z_{inr} \leq \ell_{nr} + M_{inr} y_{inr} && \forall i \in \mathcal{C}, n, r \in R_s \quad (3.73)\\
& U_{inr} - M_{inr}(1 - y_{inr}) \leq z_{inr} && \forall i \in \mathcal{C}, n, r \in R_s \quad (3.74)\\
& z_{inr} \leq U_{inr} && \forall i \in \mathcal{C}, n, r \in R_s \quad (3.75)\\
& z_{inr} \leq U_{nr} && \forall i \in \mathcal{C}, n, r \in R_s \quad (3.76)\\
& U_{nr} \leq z_{inr} + M_{inr}(1 - w_{inr}) && \forall i \in \mathcal{C}, n, r \in R_s \quad (3.77)\\
& \sum_{i\in\mathcal{C}} w_{inr} = 1 && \forall n, r \in R_s \quad (3.78)\\
& w_{inr} \leq y_{inr} && \forall i \in \mathcal{C}, n, r \in R_s \quad (3.79)\\
& \sum_{m=1}^{n} w_{imr} \leq (c_i - 1) y_{inr} + (n-1)(1 - y_{inr}) && \forall i \in \mathcal{C}, n > c_i, r \in R_s \quad (3.80)\\
& c_i(1 - y_{inr}) \leq \sum_{m=1}^{n} w_{imr} && \forall i \in \mathcal{C}, n > 1, r \in R_s \quad (3.81)\\
& a_{in} w_{inr} \leq \eta_{inr} && \forall i \in \mathcal{C}\setminus\{0\}, n, r \in R_s \quad (3.82)\\
& \eta_{inr} \leq b_{in} w_{inr} && \forall i \in \mathcal{C}\setminus\{0\}, n, r \in R_s \quad (3.83)\\
& p_{in}^{s} - (1 - w_{inr}) b_{in} \leq \eta_{inr} && \forall i \in \mathcal{C}\setminus\{0\}, n, r \in R_s \quad (3.84)\\
& \eta_{inr} \leq p_{in}^{s} - (1 - w_{inr}) a_{in} && \forall i \in \mathcal{C}\setminus\{0\}, n, r \in R_s \quad (3.85)\\
& z_{inr} \leq d_{0nr} w_{0nr} + \sum_{j\in\mathcal{C}\setminus\{0\}} (\beta_{jn}\eta_{jnr} + d_{jnr} w_{jnr}) && \forall i \in \mathcal{C}, n, r \in R_s \quad (3.86)\\
& y_{inr}, w_{inr} \in \{0, 1\} && \forall i \in \mathcal{C}, n, r \in R_s \quad (3.87)
\end{aligned}
$$

Model 3.9: Lagrangian subproblem associated with subset s for the Lagrangian decomposition on the price variables of the revenue maximization problem (Model 3.1)

3.5.2 Feasible solutions

The procedure described in Section 3.5.1 provides an upper bound of the optimal value of the objective function of the revenue maximization problem (Model 3.1) for given values of the Lagrangian multipliers, but it does not necessarily yield a feasible solution. We are interested in the generation of feasible solutions for two reasons. First, we want to determine solutions to the original problem at each iteration of the subgradient method.

Second, it enables to obtain lower bounds for the revenue maximization problem, which are given by any feasible solution (in the case of a maximization problem), in order to calculate the duality gap. This gap, which is defined as the relative difference between the upper and lower bounds, allows us to assess the quality of the feasible solution. Hence, we can set a threshold on how low the duality gap should be (based on the specific requirements of the solution) and use it as a stopping criterion in the subgradient method.

The Lagrangian solution obtained after solving the subproblems defined in Section 3.5.1 for given values of the Lagrangian multipliers is feasible for the revenue maximization problem if it satisfies the relaxed constraints, i.e., if all the price duplicates $p_{in}^s, \forall i \in \mathcal{C} \setminus \{0\}, n$, are the same across subsets. In fact, given that the dualized constraints are equality constraints, feasibility of the Lagrangian solution automatically implies optimality (Guignard, 2003).

Despite it is quite rare in practice to have feasible Lagrangian solutions, in this case we can easily derive solutions from the values of the price duplicates that are feasible for the revenue maximization problem. We denote the set of values associated with each subset s by $\{\bar{p}_{in}^s\}_s, \forall i \in \mathcal{C}_n \setminus \{0\}, n$. If the price variables in Model 3.1 are fixed to some values $\bar{p}_{in}$, the resulting problem is considerably less challenging to solve. The variables η_{inr} and the associated linearizing constraints (3.13)–(3.16) are no longer necessary because the product $\bar{p}_{in} w_{inr}$ does not need to be linearized. Furthermore, the problem can be solved by iterating over the draws and the customers while updating the availability of the services, which is similar to the solution procedure outlined in Section 3.2 (see Algorithm 3.1) when the price is proposed at the individual level. The solution approach associated with the revenue maximization problem (Model 3.1) with fixed prices $\bar{p}_{in}$ is depicted in Algorithm 3.2. For each draw we initialize the occupancy level of all services in $\mathcal{C} \setminus \{0\}$ to 0, which means that all the capacity is available (note that the opt-out option is always available so it does not need to be tracked). We then iterate over the individuals in the order provided by the priority list, and for each individual we calculate the utilities of the available services and set the choice to the service achieving the highest utility. Finally, we increase the occupancy associated with the chosen service by one unit and update the total objective function.

There exist different strategies to set values for the price variables p_{in} from the values in $\{\bar{p}_{in}^s\}_s$. We could rely, for instance, on a statistical quantity, such as the mean or the median, or the extrema across subsets of draws. These approaches determine a single price configuration for each service and customer, and do not allow to explore the extent of possibilities that can be generated by considering independently the values in $\{\bar{p}_{in}^s\}_s$. We propose a methodology that iterates over the subsets of draws and for each subset solves Model 3.1 with the prices fixed to the values of the price copies for that subset using Algorithm 3.2. This generates S feasible solutions for the revenue maximization problem, among which we select the one with the highest objective function. Algorithm 3.3 summarizes this procedure.

Algorithm 3.2: Solution approach for the revenue maximization problem (Model 3.1) with fixed prices $\bar{p}_{in}$

Input: Fixed prices $\bar{p}_{in}$, revenue maximization problem (Model 3.1);
Output: Optimal values for the remaining variables (y_{inr}, w_{inr}, U_{inr} and U_{nr}) and objective function Z;

Initialize the objective function $Z = 0$;
for $r = 1 \dots R$ **do**
  Initialize occupancy level $o_{ir} = 0 \; \forall i \in \mathcal{C} \setminus \{0\}$ and availability variables $y_{inr} = 1, \forall i \in \mathcal{C}_n \setminus \{0\}, n$;
  for $n = 1 \dots N$ **do**
    for $i \in \mathcal{C}_n \setminus \{0\}$ **do**
      if $o_{ir} < c_i$ **then**
        Calculate $U_{inr} = \beta_{in}\bar{p}_{in} + d_{inr}$;
      else
        Set the alternative unavailable: $y_{inr} = 0$;
        Set the utility to the corresponding lower bound: $U_{inr} = \ell_{nr}$;
    Calculate $U_{0nr} = d_{0nr}$;
    Determine $U_{nr} = \max_{\{i \in \mathcal{C}_n | y_{inr}=1\} \cup \{0\}} U_{inr}$ and $j = \operatorname{argmax} U_{nr}$;
    Set $w_{jnr} = 1$ and $w_{inr} = 0 \; \forall i \in \mathcal{C}_n \setminus \{j\}$;
    Update the objective function $Z = Z + \sum_{i \in \mathcal{C}_n \setminus \{0\}} \frac{1}{R} w_{inr}\bar{p}_{in}$;
    Update the occupancy level $o_{jr} = o_{jr} + 1$;

Algorithm 3.3: Computation of feasible solutions for the revenue maximization problem (Model 3.1) by iterating over the subsets of draws

Input: Set of values for the price duplicates $\{\bar{p}^s_{in}\}_s$ obtained after solving the corresponding Lagragian subproblems (Model 3.9);
Output: The feasible solution of the revenue maximization problem (Model 3.1) that provides the highest lower bound on Z among the generated feasible solutions;

Initialize the best lower bound $Z_{\mathrm{LB}} = -\infty$;
for $s = 1 \dots S$ **do**
  Set $\bar{p}_{in} = \bar{p}^s_{in} \; \forall i \in \mathcal{C}_n \setminus \{0\}, n$;
  Solve the revenue maximization problem (Model 3.1) with fixed prices $\bar{p}_{in}$ using Algorithm 3.2 and obtain current lower bound Z^s_{LB};
  if $Z^s_{LB} > Z_{LB}$ **then**
    Update Z_{LB} and keep the associated solution as the feasible solution with the highest objective function among the evaluated ones so far;

Notice that we do not determine all the price configurations that can be characterized with the set $\{\bar{p}_{in}^{s}\}_s$, as the variables p_{in} are fixed to the values of the price copies associated with a single subset at each iteration. This means that we do not allow for configurations with price copies from different subsets. Hence, this procedure represents a simplification of a brute force search where all possible candidates are evaluated. Even though Algorithm 3.2 provides a straightforward solution approach for Model 3.1 when the prices are fixed, trying all possible combinations might induce a notable computational burden, especially in the case of large instances (i.e., large number of draws and customers) and/or a small size of the subsets of draws.

Section 3.5.3 explains the subgradient method. It integrates the Lagrangian subproblem detailed in Section 3.5.1, which provides an upper bound for the revenue maximization problem (Model 3.1), and the computation of feasible solutions via Algorithm 3.3, which generates a lower bound.

3.5.3 Subgradient method

The tightest possible bound obtained from the upper bounds $Z^{\text{UB}}(\alpha)$ for the revenue maximization problem (Model 3.1) as a function of the Lagrangian multipliers α is attained by solving the Lagrangian dual:

$$Z^{\text{LD}} = \min_{\alpha} Z^{\text{UB}}(\alpha). \tag{3.88}$$

The Lagrangian subproblem (Model 3.9) does not have the integrality property because there might be optimal solutions to its LP relaxation that are not integer. As a result, the Lagrangian decomposition scheme yields stronger bounds than the LP relaxation, i.e., $Z \leq Z^{\text{LD}} \leq Z^{\text{LP}}$, where Z^{LP} denotes the optimal value of the objective function of the LP relaxation of Model 3.1.

The subgradient method is an iterative procedure that can be employed to solve an optimization problem whose objective function is a non-differentiable convex function on a closed convex set. This is our case, as the objective function of the Lagrangian dual is non-differentiable and convex (it is the upper envelope of a finite family of linear functions) and $\mathbb{R}$ is the set of admissible values for the Lagrangian multipliers. This method constructs a sequence $\{\alpha^k\}_k$ using

$$\alpha^{k+1} = \alpha^k + \gamma^k v^k, \qquad \forall k, \tag{3.89}$$

where k denotes the iteration, γ^k a positive scalar called step size and v^k a vector representing the direction of motion called step direction. Algorithm 3.4 presents the subgradient method. A detailed explanation is provided in the following.

Algorithm 3.4: Subgradient method

Input: Limit on the computational time T and number of iterations ω for the update of λ^k;

Output: $Z^{\text{UB,best}}$, $Z^{\text{LB,best}}$ and a feasible solution for the revenue maximization problem (Model 3.1);

Initialize $k = 0$, $t = 0$, $Z^{\text{UB,best}} = +\infty$, $Z^{\text{LB,best}} = -\infty$, λ^0, $\alpha^0_{ins} = 0$ and $v^0_{ins} = 0$ $\forall i \in \mathcal{C}_n \setminus \{0\}, n, s$;

while $t < T$ **do**

  for $s = 1 \dots S$ **do**

    Solve the Lagrangian subproblem (Model 3.9) and obtain $\{\bar{p}^s_{in}\}_s$ and $Z^{\text{UB}}(\alpha^k)$;

  if $Z^{UB}(\alpha^k) < Z^{UB,best}$ **then**

    Update the best upper bound found so far $Z^{\text{UB,best}} = Z^{\text{UB}}(\alpha^k)$;

  Compute the feasible solution providing the largest lower bound using Algorithm 3.3 and obtain Z^{LB};

  if $Z^{LB} > Z^{LB,best}$ **then**

    Update the best lower bound found so far $Z^{\text{LB,best}} = Z^{\text{LB}}$;

  if $Z^{UB}(\alpha^k)$ *has not improved in the last* ω *consecutive iterations* **then**

    Set $\lambda^k \leftarrow \lambda^k/2$;

  Calculate γ^k according to (3.93);

  Update the Lagrangian multipliers $\alpha^{k+1}_{ins} = \alpha^k_{ins} + \gamma^k v^k_{ins}$, where $v^k_{ins} = -(g^k_{ins} + \zeta^k v^{k-1}_{ins})$ and ζ^k is calculated according to (3.92);

  Update the computational time t;

  Set $k \leftarrow k + 1$;

A subgradient of $Z^{\text{UB}}(\alpha)$ at $\bar{\alpha}$ is a multidimensional vector g of dimension $D = (|\mathcal{C}| - 1)NS$ that satisfies

$$Z^{\text{UB}}(\alpha) \geq Z^{\text{UB}}(\bar{\alpha}) + g(\alpha - \bar{\alpha}), \qquad \forall \bar{\alpha} \in \mathbb{R}^D. \tag{3.90}$$

The vector g defined as $g_{ins} = p^s_{in} - p^{(s+1)}_{in}$, $\forall i \in \mathcal{C} \setminus \{0\}, n, s$ and evaluated at the optimal solution of the Lagrangian subproblems is a subgradient of $Z^{\text{UB}}(\alpha)$ at any admissible α (Fisher, 1981b).

The Lagrangian multipliers are initialized to 0, i.e., $\alpha^k_{ins} = 0, \forall i \in \mathcal{C} \setminus \{0\}, n, s$, and are updated at each iteration $k \geq 1$ by taking a step size γ^k in the direction v^k_{ins}. We take a step in the direction of the negative subgradient (the problem to be optimized here is the Lagrangian dual, which is a minimization problem).

Some tests performed on small instances show that the angle between the current step direction v^k_{ins} and the previous one v^{k-1}_{ins} is obtuse in multiple occasions. This leads to a next iterate that is close to the previous multiplier, which slows down the convergence of the procedure. This effect is known as zigzagging of kind I, and can be overcome by deflecting the step direction (Camerini et al., 1975). Thus, we define the direction of motion at iteration k as

$$v^k = -(g^k + \zeta^k v^{k-1}), \tag{3.91}$$

where $\zeta^k \in \mathbb{R}_{\geq 0}$ is a suitable scalar called deflection parameter. By defining this parameter as

$$\zeta^k = \begin{cases} -\tau \frac{g^k v^{k-1}}{\|v^{k-1}\|^2} & \text{if } g^k v^{k-1} < 0, \\ 0 & \text{otherwise,} \end{cases} \tag{3.92}$$

with $1 \leq \tau < 2$, the step direction is forced to always form an acute angle with the preceding one, which eliminates the zigzagging of kind I. The use of $\tau = 1.5$ is recommended.

The step size most commonly used in practice is defined by Held et al. (1974) for each iteration k as follows:

$$\gamma^k = \lambda^k \frac{Z^{\text{UB}}(\alpha^k) - Z}{\|v^k\|^2}, \tag{3.93}$$

where $Z^{\text{UB}}(\alpha^k)$ is the objective function of the Lagrangian subproblem for the Lagrangian multipliers α^k, Z is the optimal objective function in Model 3.1, which can be replaced by the best lower bound found so far, $\|v^k\|^2$ is the norm of the step direction, i.e., $\|v^k\|^2 = \sum_{i \in \mathcal{C} \setminus \{0\}} \sum_{n=1}^{N} \sum_{s=1}^{S} (v^k_{ins})^2$, and λ^k is a step size decreasing parameter satisfying $0 < \lambda^k \leq 2$. As described in Fisher (1973), this value can be initially set to $\lambda^0 = 2$ and halved whenever $Z^{\text{UB}}(\alpha^k)$ has failed to decrease in some fixed number ω of consecutive iterations.

Notice that unlike the ordinary gradient method, the subgradient method is not a descent method, as the subgradient might not be a descent direction. Furthermore, even when it is a descent direction, the step size can be such that the next iterate provides a larger objective function. This is why we keep track of the best upper and lower bounds found throughout the method.

With respect to the termination of the method, Algorithm 3.4 defines a stopping criteria on the computational time, as is considered in the experiments performed in Section 3.6. This allows us to evaluate the computational performance with respect to the exact method and the trade-off between the different dimensions the Lagrangian subproblem is built on (namely, customers, draws and subsets of draws). Other stopping criteria could be considered, such as a maximum number of iterations, or a certain number of iterations after which the upper bound has not experienced an improvement.

In summary, the subgradient method approximates the Lagrangian dual by updating the values of the Lagrangian multipliers at every iteration. Provided initial values of the multipliers, it produces new multipliers by applying to the previous ones a direction of motion, which is calculated in terms of the subgradient of the objective function of the Lagrangian subproblem, and a step size. As a result, we obtain a set of multipliers that are employed in the next iteration to solve the Lagrangian subproblems. We also benefit from the generation of feasible solutions in the update of the multipliers. Indeed, the calculation of the step size here considered requires the optimal value of the objective function of the problem being relaxed, which is typically replaced by the best lower bound found so far (in a maximization problem).

3.6 Case study

The goal of this case study is to test the computational efficiency and quality of the obtained solutions of the Lagrangian decomposition scheme developed in Section 3.5. Furthermore, we also provide an assessment of the trade-off between these two elements and the three dimensions of the approach, namely the customers (N), the draws (R) and the size of the subsets of draws (S).

To this end, we rely on the parking choices case study described in Section 2.4. In Ibeas et al. (2014), 197 individuals are interviewed in order to be able to model their preferences with respect to three parking services: paid on-street parking (PSP), paid parking in an underground car park (PUP) and free on-street parking (FSP), which represents the opt-out option. With the collected data, a mixtures of logit model is specified and estimated. As is the case for the error term ε_{in} of the utility function, the parameters of the choice model that are assumed to follow a probability distribution to capture individuals' heterogeneity are drawn prior to solving the problem in an exact manner or applying the decomposition scheme. Likewise, we assume that the same price is offered to all customers, i.e., $p_{in} = p_i, \forall n$, not only because it is more appropriate in a parking context, but also because it yields a problem that is more aggregated, in the sense that a decomposition by individual or group of individuals cannot be applied.

The instances considered to evaluate the approach are defined from the available data by setting specific values for N, R and S. For a given number of customers N, the capacity of the services is defined accordingly such that it is appropriate for the size of the population but restrictive enough so that some customers are forced to choose the opt-out option because some of the services might become unavailable. Similarly to Section 2.4, for each configuration determined by N, R and S, we define three different instances by randomly selecting N individuals from the whole dataset, together with the associated draws of the random parameters and the error term. Notice that the instances with $N = 197$ contain the same individuals but might have a different priority list, i.e., the individuals might be processed in a different order.

Section 3.6.1 analyzes the performance of the decomposition framework with respect to the exact method, and given the complexity of the latter, we consider a single instance for each configuration of N, R and S being tested. Section 3.6.2 assesses the computational complexity and quality of the obtained bounds with respect to the three above-mentioned dimensions by fixing the values for two of them at a time and varying the remaining one.

The calibration of the parameters of the subgradient method has been performed on different configurations of N, R and S. We consider $\lambda^0 = 0.5$ because larger values result in a higher increase of the upper bound in the first iterations, and $\omega = 5$ because in some experiments a low number of iterations is expected. We follow the recommendations found in the literature and we set $\tau = 1.5$. The developed computer codes have been implemented in C++ using ILOG Concert Technology to access CPLEX 12.8, and all the instances were performed using 12 threads in a 3.33 GHz Intel Xeon X5680 server running a 64-bit Ubuntu 16.04.2.

3.6.1 Comparison with optimal solutions

In this section, we compare the best feasible solution obtained from the Lagrangian decomposition scheme with the optimal solution for three configurations consisting of $N = 50$, $R \in \{100, 250, 500\}$ and $S = 5$, with a capacity of 20 customers for both PSP and PUP ($c_{\text{PSP}} = c_{\text{PUP}} = 20$). Given the computational complexity associated with solving exactly the MILP formulation, we set restrained values for N, R and S, the latter being selected in relation to the number of customers so that the associated subproblems are computationally not too complex. Moreover, we run only one instance for each configuration. For the sake of the experiment, we set a maximum running time T equal to 10% of the computational time of the exact approach of each instance to run the Lagrangian decomposition scheme.

Table 3.1 provides an overview of the performance of both approaches with respect to computational efficiency and quality of the obtained solutions. For the exact method, it includes the computational time and the optimal value of the objective function. For the Lagrangian decomposition scheme, it reports the number of iterations K achieved within the established time limit T (10% of the computational time reported by the exact method), the best upper and lower bounds found together with the iteration at which they were reached (in brackets), the average computational time per iteration $\bar{t}_k$, the duality gap (gap_{dual}) and the gap between the best feasible solution and the optimal one (gap_{opt}). These gaps are calculated as follows:

$$\text{gap}_{\text{opt}} = \frac{Z - Z^{\text{LB,best}}}{Z}, \tag{3.94}$$

$$\text{gap}_{\text{dual}} = \frac{Z^{\text{UB,best}} - Z^{\text{LB,best}}}{Z^{\text{LB,best}}}. \tag{3.95}$$

The number of iterations performed within the established computational time limit T grows as R increases. As shown by the number in brackets in columns $Z^{\mathrm{UB,best}}$ and $Z^{\mathrm{LB,best}}$, the best lower bound, which provides a feasible solution, is found relatively soon with respect to the total number of performed iterations K, as opposed to the best upper bound, which keeps on improving. The relative difference between bounds is in all instances lower than 2%, and the gap with respect to the optimal solution is way smaller, with a solution almost equivalent to the optimal one for the largest tested instance.

Table 3.1: Comparison of the performance of the exact method and the Lagrangian decomposition scheme for $N = 50$, $R \in \{100, 250, 500\}$ and $S = 5$

	Exact method		Lagrangian decomposition scheme					
R	Time	Z	K	$Z^{\mathrm{UB,best}}$ (k)	$Z^{\mathrm{LB,best}}$ (k)	$\bar{t}_k$(min)	$\mathrm{gap}_{\mathrm{dual}}$(%)	$\mathrm{gap}_{\mathrm{opt}}$(%)
100	3.76 h	26.21	5	26.70 (5)	26.18 (2)	5.16	1.98	0.11
250	36.3 h	26.04	14	26.46 (14)	26.02 (1)	16.8	1.70	0.09
500	5.07 days	25.99	21	26.40 (21)	25.99 (7)	35.8	1.58	0.02

Figure 3.2 portrays the evolution of the upper and lower bounds (Z^{UB} and Z^{LB}) throughout the iterations of the subgradient method. In all cases, the upper bound experiences an increase during the first iterations followed by a decreasing trend that tends to stabilize. For $R = 250$ and $R = 500$, we observe such stabilization from (approximately) $k = 10$. The lower bound, instead, remains much closer to the optimal value of the objective function (Z), already from the early iterations, and it practically does not fluctuate, achieving the same values in more than one iteration. This means that the same price configuration is obtained in different iterations and it reports the highest revenue at each of them.

We can also compare the prices and expected demand of the optimal solution with respect to the best feasible solution found in the Lagrangian decomposition scheme. As shown in Table 3.2, the price of PSP of the best feasible solution is slightly closer to the corresponding optimal value than the price of PUP. Modest differences are as well observed when comparing the expected demand of both solutions, with a tendency for the best feasible solution to generate larger values of the expected demand of FSP.

Table 3.2: Comparison of the prices and the expected demand of the optimal solution (Ex.) and the best feasible solution (Dec.) found in the Lagrangian decomposition scheme for $N = 50$, $R \in \{100, 250, 500\}$ and $S = 5$

	p_{PSP}		p_{PUP}		D_{PSP}		D_{PUP}		D_{FSP}	
R	Ex.	Dec.	Ex.	Dec.	Ex.	Dec.	Ex.	Dec.	Ex.	Dec.
100	0.609	0.610	0.840	0.837	19.07	18.99	17.39	17.45	13.54	13.56
250	0.591	0.591	0.793	0.811	19.18	19.46	18.56	17.90	12.26	12.64
500	0.588	0.591	0.790	0.792	19.29	19.56	18.55	17.24	12.16	13.20

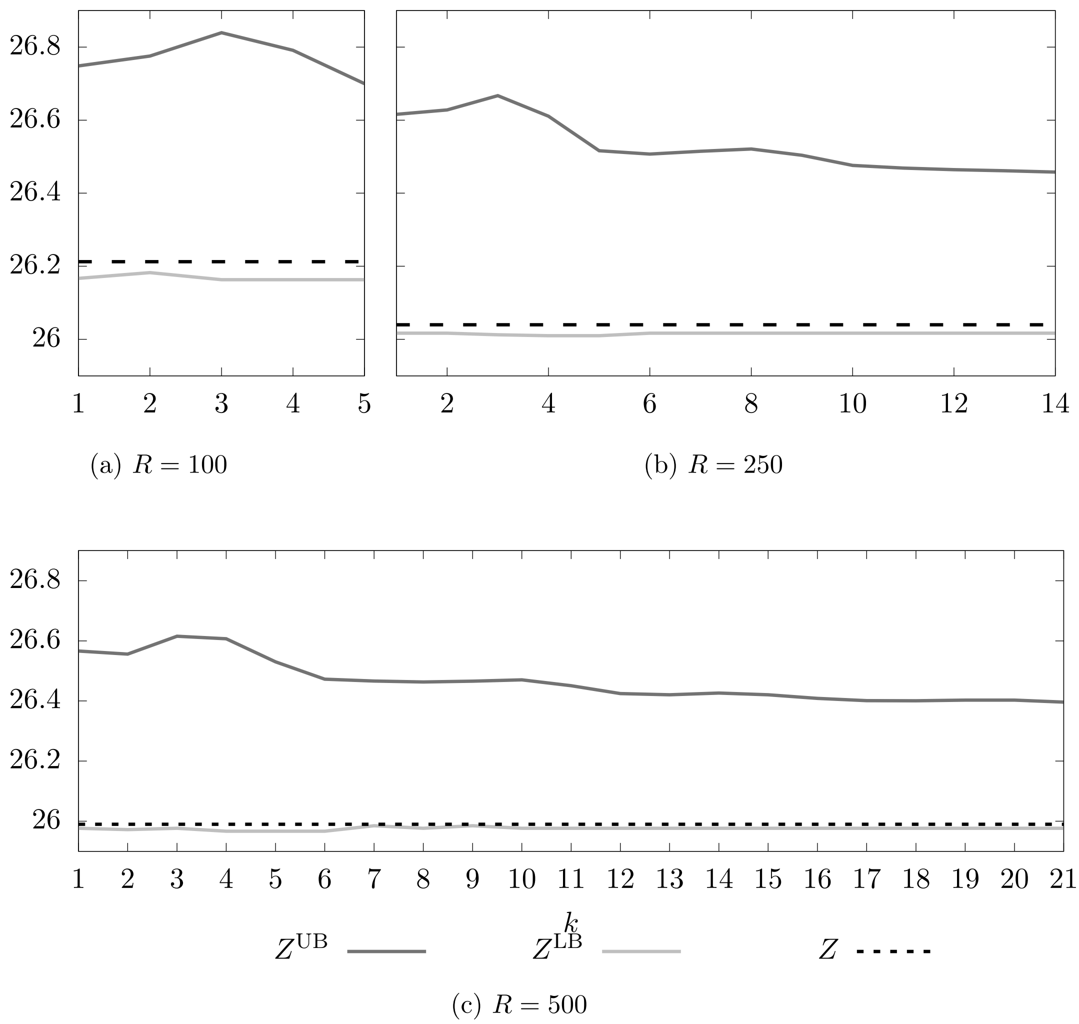

Figure 3.2: Evolution of the lower and upper bounds throughout the iterations of the Lagrangian decomposition scheme for $N = 50$, $R \in \{100, 250, 500\}$ and $S = 5$

3.6.2 Trade-off between R, N and S

We evaluate now the computational complexity and quality of the obtained bounds by setting a base configuration for N (number of individuals), R (number of simulation draws) and S (size of the subsets of draws) and varying one dimension at a time. More precisely, the base configuration is defined by $N = 50$ as in Section 2.4, $R = 500$ to ensure precision of the results, and $S = 2$ to deal with small subproblems so that the algorithm is able to run multiple iterations within the established time limit. Recall that for each configuration we run three different instances defined by a random selection of N individuals from the whole dataset with the associated draws for the error term and random parameters. For the sake of evaluating the performance of the algorithm, we set a large enough computational time limit T for each configuration such that the algorithm is able to run for some iterations.

Variation of R

In this experiment, we allow for values of $R \in \{100, 250, 500, 1000, 2500, 5000\}$. Table 3.3 includes the computational time limit T, the average time per iteration associated with the computation of the upper and lower bound ($\bar{t}_{k,\mathrm{UB}}$ and $\bar{t}_{k,\mathrm{LB}}$, respectively), and the average number of iterations performed within T (denoted by $\bar{K}$).

We observe that the growing in $\bar{t}_{k,\mathrm{UB}}$ with respect to R is more moderate than the exponential growing experienced with the exact method, and although $\bar{t}_{k,\mathrm{LB}}$ does not present a similar growth rate, it corresponds to an insignificant fraction of the entire computational time of an iteration. Nevertheless, notice that this would not be the case if a brute force search was applied, as the resulting computational time per iteration would be of the order of $(R/S) \cdot \bar{t}_{k,\mathrm{LB}}$.

Table 3.3: Computational performance of the Lagrangian decomposition scheme for the 3 considered instances defined by $N = 50$, $S = 2$ and $R \in \{100, 250, 500, 1000, 2500, 5000\}$

R	T (min)	$\bar{t}_{k,\mathrm{UB}}$ (min)	$\bar{t}_{k,\mathrm{LB}}$ (s)	$\bar{K}$
100	30	1.33	0.013	23.3
250	75	3.67	0.078	21.3
500	150	7.91	0.32	19.3
1000	300	18.4	1.72	17
2500	750	57.1	14.6	13.7
5000	1500	145.4	62.1	11.7

Figure 3.3 exhibits $\mathrm{gap}_{\mathrm{dual}}$ for the three instances associated with each value of R, as well as the iteration at which the best lower bound (i.e., the best feasible solution) is achieved (in brackets). We notice an increasing trend for somewhat larger gaps as R increases, which is related to the fact that the increase in time to solve the Lagrangian subproblems is not linear, and therefore an increase in T would be needed to reach gaps of the same order as the ones that are reached for lower values of R. We also observe that the gaps are larger than the ones achieved in the previous experiment because we have considered here $S = 2$ instead of $S = 5$ (see Section 3.6.2 for an extensive assessment on S).

As already seen in Section 3.6.1, the best lower bound is generally obtained during the first iterations of the subgradient method, with a couple of exceptions for the largest values of R. Hence, it would be possible to set a threshold on the number of iterations without improvement of the best lower bound and use it as a stopping criterion of the algorithm (when the duality gap is below an acceptable threshold).

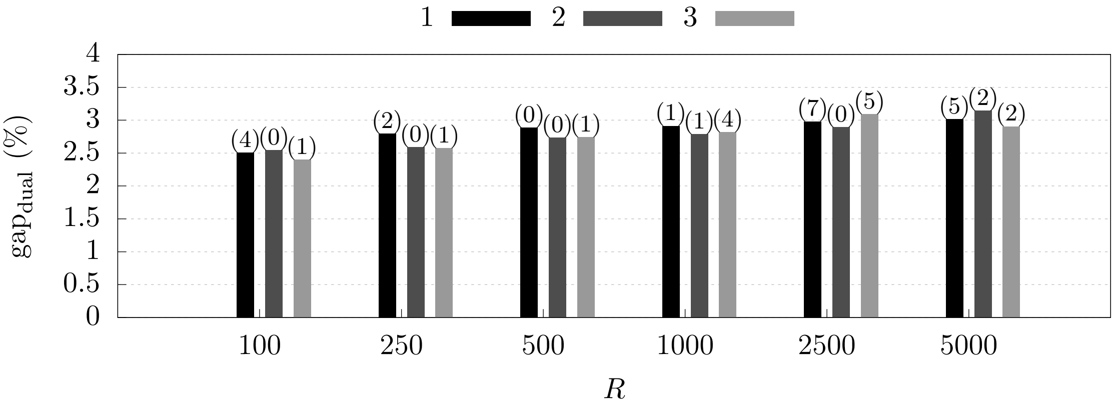

Figure 3.3: Duality gap (gap_{dual}) and iteration at which the best lower bound is achieved (in brackets) for the 3 considered instances (1-3) for $R \in \{100, 250, 500, 1000, 2500, 5000\}$

Variation of N

The same analysis can be carried out with respect to the size of the population under consideration. We consider $N \in \{50, 100, 150, 197\}$. Notice that in this case the capacity of the services needs to be defined accordingly. As discussed in Section 2.4, a capacity $c_{\text{PSP}} = c_{\text{PSP}} = 20$ spots for $N = 50$ is assumed for both services because it is large enough to be realistic for the size of the sample but restrictive enough to force some individuals to opt-out. Based on the same idea, we determine the capacity of the services for larger values of N.

As presented in Table 3.4, the average time to solve the Lagrangian subproblems ($\bar{t}_{k,\text{UB}}$) has an exponential growing. In the case of $N = 150$ and $N = 197$, we even obtain larger values than the computational time limit T, which is due to the fact that the first iteration is always performed and the stopping criterion on the computational time is verified from $k = 1$.

Table 3.4: Computational performance of the Lagrangian decomposition scheme for the 3 considered instances defined by $R = 500$, $S = 2$ and $N \in \{50, 100, 150, 197\}$

N	$c_{\text{PSP}} = c_{\text{PUP}}$	T (min)	$\bar{t}_{k,\text{UB}}$ (min)	$\bar{t}_{k,\text{LB}}$ (s)	$\bar{K}$
50	20	150	7.91	0.32	19.3
100	40	300	95.1	0.78	3.3
150	60	450	460.7	1.54	1.3
197	80	600	1088.5	2.39	1

As shown in Figure 3.4, the duality gap presents an opposite trend than the previous experiment, since it is inversely proportional to N. Moreover, we also observe an earlier attainment of the best lower bound for low values of N. Hence, although less iterations are performed as N increases, it does not appear to be a drawback because lower duality gaps are reached.

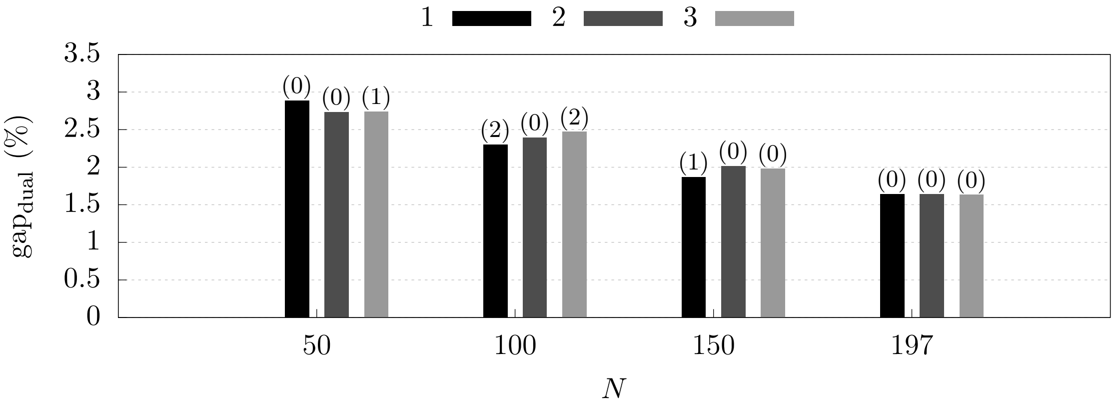

Figure 3.4: Duality gap (gap$_{\text{dual}}$) and iteration at which the best lower bound is achieved (in brackets) for the 3 considered instances (1-3) for $N \in \{50, 100, 150, 197\}$

Variation of S

For this experiment, we set the same computational time limit T for $S \in \{1, 2, 3, 4, 5, 10\}$. The idea is to evaluate the trade-off between the size of the subsets of draws and the quality of the gap obtained for each subset within the established time limit. As expected, the larger the size of the subsets, the higher the computational time per iteration, and the lower the total number of iterations (see Table 3.5). Furthermore, since the number of price configurations to be tested decreases with the increase in the subset size, the computational time dedicated to find a feasible solution diminishes as S increases.

Table 3.5: Computational performance of the Lagrangian decomposition scheme for the 3 considered instances defined by $N = 50$, $R = 500$ and $S \in \{1, 2, 3, 4, 5, 10\}$

S	T (min)	$\bar{t}_{k,\text{UB}}$ (min)	$\bar{t}_{k,\text{LB}}$ (s)	$\bar{K}$
1	150	2.41	0.672	62.7
2	150	7.91	0.323	19.3
3	150	12.4	0.213	12.7
4	150	21.5	0.158	7.7
5	150	46.1	0.136	4
10	150	126.9	0.065	2

With respect to the gap, Figure 3.5 depicts its decrease with respect to the increase in S. Such decrease is specially notable from $S = 1$ to $S = 2$, where the gap goes from values that oscillate between 3.46% to 3.74% to 2.73% to 2.89%. This confirms our expectations about the restricted potential of the decomposition by single draws, which has completed in average 62.7 iterations but has not managed to lessen the gap or to improve the best feasible solution after the first iterations.

We observe a moderate decrease in the gaps from $S = 2$ to $S = 3$, and similar values for $S = 4$ and $S = 5$. The largest size, $S = 10$, reports the best gaps, with values

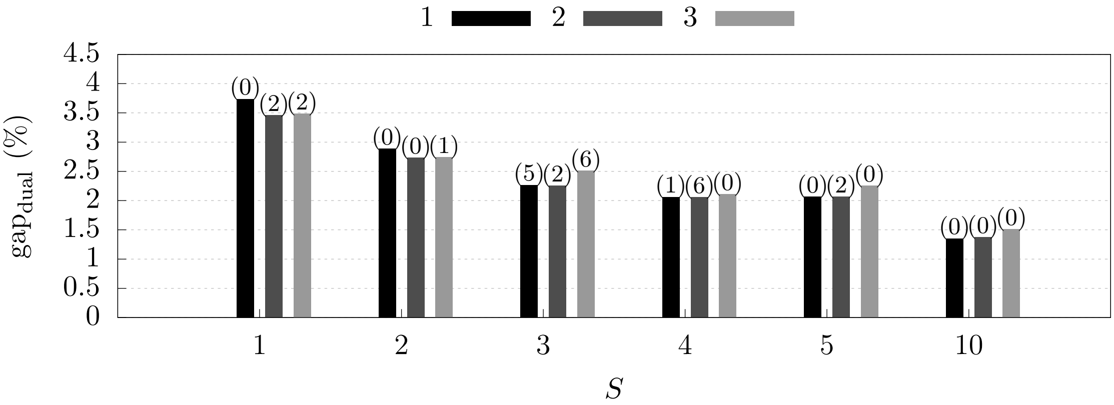

Figure 3.5: Duality gap (gap_{dual}) and iteration at which the best lower bound is achieved (in brackets) for the 3 considered instances (1-3) for $S \in \{1, 2, 3, 4, 5, 10\}$

between 1.35% and 1.51%. Hence, even though the algorithm runs for a limited number of iterations ($K = 2$ for the three instances), it is beneficial to consider large subsets of draws in order to improve the quality of the obtained bounds for a given running time. Notice that in this case the total running time exceeds the time limit, as the stopping criterion allows for an additional iteration if the time limit has not been reached.

Nevertheless, we cannot provide a generalization for the appropriate value of S, as it depends on the number of customers. Indeed, $S = 10$ might be computationally too complex for higher values of N (e.g., $N = 197$), which implies that we might need to rely on lower sizes of the subsets for the algorithm to be operational. As we have noticed in the experiment that varies the number of individuals N, the fact that a low value of S is assumed for larger values of N does not necessarily imply a decrease in the quality of the obtained bounds.

3.7 Concluding remarks

In this chapter, we have provided a comprehensive literature review on decomposition techniques applied to mixed-integer linear optimization. In the light of the main findings, we have initially developed a strategy that identifies two interesting subproblems (one associated with the operator and one associated with the customers) and separates them by relaxing the linking constraint. Due to the strong interactions between both subproblems, such decomposition results in a weak relaxation. Thus, we decided to rely on Lagrangian decomposition to induce separability while preserving all the constraints within the resulting subproblems. This decomposition scheme is then generalized to address some of its limitations by gathering simulation draws into subsets and generating copies of the variables being split for each subset, as opposed to the original attempt where a copy is introduced for each draw. The solving of the Lagrangian subproblems, together with the computation of feasible solutions from the Lagrangian solution, is enclosed in the well-known subgradient method.

As is the case in Chapter 2, we specify the Lagrangian decomposition scheme in terms of the revenue maximization problem described in Section 3.2, which considers the price as the only endogenous variable and performs some assumptions such as fixed capacity and availability at the operator level to all customers. Nevertheless, we notice that the learning acquired from the initial attempts, as well as the proposed methodology, can be applied to other problems with different endogenous variables and assumptions.

The experiments performed on the parking case study described in Section 2.4 show that near-optimal solutions can be obtained in a much lower computational time with respect to the exact approach. Such solutions are usually found at an early stage of the subgradient method, which enables to terminate the algorithm when the best feasible solution found by the method has not improved after a certain number of consecutive iterations and the reported duality gap is below an acceptable threshold. Furthermore, we have seen that as long as the corresponding subproblems are computationally manageable, a large number of draws per subset is recommended, as it leads to smaller duality gaps for a given running time.

Notice that it is possible to reduce the total computational time of the decompositon framework by implementing a parallelization routine within the subgradient algorithm that allows to solve multiple subproblems simultaneously. The values of the price copies obtained after solving the subproblems need to be gathered so that the subgradients can be calculated at the end of each iteration of the subgradient method. We can also apply parallelization to the computation of feasible solutions, but as shown in Section 3.6, the associated computational burden is negligible with respect to the solving of the subproblems, so it will have a mild impact on the total running time.

4

Welfare-maximizing design of a transportation system

Preliminary ideas related to this chapter are included in the conference paper

Pacheco, M., Sharif Azadeh, S., and Bierlaire, M. (2019). Passenger satisfaction maximization within a demand-based optimization framework. In *19th Swiss Transport Research Conference (STRC), Ascona, Switzerland*

The work has been performed by the candidate under the supervision of Prof. Shadi Sharif Azadeh and Prof. Michel Bierlaire.

4.1 Introduction

The choice-based optimization framework detailed in Chapter 2 allows to incorporate a disaggregate demand representation into an MILP model. In Chapters 2 and 3, we illustrated the idea with the example of an operator that aims at maximizing its revenue. In this chapter, we show that the framework can accommodate other types of optimization problems. We consider here the point of view of public authorities, who want to invest in transportation projects to improve the social and economic welfare of the population, or to decrease various negative externalities.

One such relevant application can be found in the context of road pricing. According to the theory of welfare economics and externalities (Pigou, 1920), a tax or toll is needed to correct the negative externalities associated with urban road transportation, such as green-house gas emissions, air pollutants and congestion. Road pricing can be seen as a supplementary economic tool to achieve the social welfare-maximum benefit or system-minimum overall cost of passenger transportation in a network. In practice, policy instruments associated with road pricing include mostly congestion pricing in the form of a toll to confront drivers with the cost of the congestion delays they impose on other drivers.

A key factor for a successful implementation of a road pricing policy is its acceptability (Kidokoro, 2010). Indeed, lack of public acceptance has been the most important barrier to these pricing schemes. Acceptability improves significantly when the availability and quality of existing public transportation (PT) service is good and the revenue raised by road pricing is recycled (Anas and Lindsey, 2011). This means that the revenue is allocated to the implementation of the system and to additional road and/or PT investments or subsidies to encourage modal shift (Lyons et al., 2004). As is typically found, in the absence of revenue recycling, the analyzed measures indicate that the road pricing scheme is regressive, i.e., the welfare is distributed more unequally after its introduction than before (Levinson, 2010).

Hence, a common requirement for the wide acceptance of a road pricing policy is to be progressive, which implies that the welfare is distributed from richer to poorer groups of society. It is therefore essential to implement welfare measures and distributional analyses to evaluate the nature of a policy. Despite their relevance, neither the change in total welfare nor the change in consumer surplus for different groups in the population are usually reported in the literature. This complicates or even prevents the use of the obtained results for public policy because it hinders the comparison of different schemes and combinations thereof (Basso and Silva, 2014).

Furthermore, we need to take into account not only the design of the road pricing scheme and its impacts on PT, but also the operation and planning of the latter as part of a comprehensive strategy of the transportation system. The studies that consider the problem of road pricing jointly with the management of PT in an economic sense generally concentrate on the effects of such interplay on system efficiency, PT operator's profitability and social welfare. In addition to fare policies, other organizational aspects of PT involved in this interplay, such as scheduling and investment strategies, might be included.

In the context of discrete choice models (DCM) based on the random utility principle, the typical measure of consumer surplus is the expected maximum utility (Ben-Akiva and Lerman, 1985). Although it has a relatively tractable form for the logit model, it is non-linear. More advanced DCM, which aim at relaxing the unrealistic assumptions associated with the logit model and have shown a better prediction power, present an even more complex formulation.

The expected maximum utility is readily available from the choice-based optimization framework thanks to the variables that capture the maximum discounted utility associated with individual n and scenario r (see definition (2.22)). This is a significant advantage because it prevents the use of the complex non-linear expressions of this quantity as provided by DCM. Hence, a linear formulation for the consumer surplus facilitates the derivation of a generalized objective function of some measure of social welfare as the policy objective, which can be employed to assess the performance of the urban transportation policies under consideration. Additionally, the disaggregate nature of the formulation provides a great deal of flexibility when it comes to define groups for distributional analyses.

We propose a model where the social welfare is to be maximized by a transportation authority that is willing to implement a road pricing scheme at the same time that the pricing and design of PT are to be adjusted in the presence of revenue recycling. The decision variables of the transportation authority are endogenous variables of the problem, as they have a direct impact on the modal split. This is a similar objective as Basso and Silva (2014), who propose a non-linear formulation for that problem. In this chapter, we illustrate the formulation using two case studies. First, we propose a semi-synthetic case study that aims at introducing a road toll in the urban region of Lausanne-Morges. Second, we translate the non-linear case study of Basso and Silva (2014) in our framework, in order to emphasize its strengths and limitations.

The remainder of the chapter is organized as follows. Section 4.2 reviews the literature on the modeling of road pricing and its interactions with PT. Section 4.3 describes how the problem of interest is accommodated in the choice-based optimization framework. Section 4.4 presents the developed case study and the numerical experiments conducted as a proof-of-concept for the introduced methodology. Section 4.5 provides an overview of the model presented in Basso and Silva (2014), the linearization scheme based on the proposed methodology and the comparison of the results obtained with both approaches. We finish the chapter with some concluding remarks in Section 4.6.

4.2 Literature review

Road pricing is an economic measure that consists of charging road users for using the road infrastructure. It provides a tool for achieving a more efficient usage of the road capacity, affecting travelers' choices of departure time, route and travel mode (among others), reducing negative external effects of road traffic (e.g., congestion, emissions), and raising revenues for funding transportation investment, operations and maintenance costs. As congestion is the most costly urban road transportation externality, and congestion relief has been the primary objective of urban road pricing (Anas and Lindsey, 2011), a vast portion of the literature in the context of road pricing is dedicated to congestion pricing.

The design of a road pricing scheme, such as congestion charging, toll cordons and toll lanes, involves decisions on when, where and how much to charge road users. Moreover, it encompasses a mathematical representation of the transportation network and demand. Most of the theoretical and practical considerations of such schemes are based on the assumptions of time-uniform charges with time-invariant (fixed or elastic) demand and travel costs (Tsekeris and Voß, 2009).

When there are no restrictions on toll locations and/or toll levels, a first-best pricing paradigm can be achieved. Such paradigm is grounded on the concept of marginal cost (Pigou, 1920, Knight, 1924), and corresponds to the unconstrained maximization of the social welfare or surplus. In reality, however, tolls can be implemented only on a subset of links of the network and/or do not include the effect of several other restrictions and market distortions. This results in second-best pricing settings (Rouwendal and Verhoef, 2006), and optimal prices deviate from marginal costs. The associated optimization problems are in general more difficult to solve than first-best problems. While the latter can be formulated as a convex optimization problem, the former is non-convex and its objective function is not everywhere differentiable.

Whereas road pricing has been extensively analyzed in the literature, its implications for PT are only partially explored (Tsekeris and Voss, 2010). The relation between both is usually examined as a stimulus-answer relation rather than an aspect of a global strategy of the transportation system. The impacts of road pricing on PT can be either direct, via changes in the generalized travel cost of private motorized modes (PMM) and PT, or indirect, which correspond to new land use patterns and transport supply that might alter structural characteristics of daily travel.

Existing methodological contributions integrating the problem of road (especially, congestion) pricing with PT strategies are commonly isolated to specific economic and operational measures. More precisely, there are studies on socially optimal combinations of road tolls with PT fares (e.g., Huang, 2000, Huang, 2002), and/or frequencies (e.g., Shepherd et al., 2006, Basso and Jara-Díaz, 2012), and subsidization (e.g., De Borger and Swysen, 1999). Other relevant aspects encompass the representation of the demand to incorporate modal split effects, changes in multi-modal network equilibrium conditions and revenue recycling.

The diversion of some users from PMM to PT will affect the modal split in favor of the latter mode. Consequently, the increase in the share of PT users might have a negative impact on the quality of the service, while partial relief of road congestion might encourage PMM users to use again their private modes. Disaggregate demand models enable to capture the relation between modal split and generalized travel cost components, which commonly include travel time and monetary costs. DCM can provide valuable insight into the feasibility evaluation of road pricing and PT management schemes, and the trade-offs between efficiency and equity impacts thereof. Other travel demand models, such as the heuristic acceptability model (Schade and Schlag, 2003) and the strategic niche management approach (Ieromonachou et al., 2007), typically rely on a priori restrictions on the demand elasticities and inter-modal substitutability, which are not underpinned by the economic theory of demand. Therefore, they cannot offer a sound theoretical basis for designing pricing policies that involve both PMM and PT users.

The effect of road pricing on PMM and PT demand has been traditionally modeled through static traffic network equilibrium models with different classes of PMM users (multi-class) and/or different modes (multi-modal), which can be combined with assumptions about the stochasticity of multiclass user equilibrium conditions and demand elasticity (e.g., Bellei et al., 2002, Hamdouch et al., 2007). However, the assumptions involved in such models do not allow to explicitly incorporate dynamic features such as traffic congestion phenomena and demand adjustment mechanisms. A small number of studies have analyzed the impact of time-varying congestion tolls (e.g., de Palma et al., 2005). Nevertheless, these approaches are usually tested on small and simplified networks, which implies that several issues associated with the deployment within large scale networks have not been properly addressed yet.

The revenues raised from road pricing can be reinvested in the transportation network to enhance the quality of PT and/or to increase the capacity of the road network, or even in other economic sectors, such as in the form of labor subsidies (e.g., Parry and Bento, 2001). However, as pointed out by Anas and Lindsey (2011), on the majority of the world's toll roads, the collected revenues are used either to cover maintenance and amortize construction costs or to make a profit for private operators, and the use of road pricing as a tool to manage demand is relatively rare. In any case, the reinvestment of revenues might involve several further considerations based on the competition regime and ownership structure of the PT market.

Early works that simultaneously address the problem of optimal road pricing and PT service design in an urban setting can be found in a bi-modal context (car and bus) for the evaluation of dedicated bus lanes in a static framework. In Mohring (1979), the optimization models developed do not yield explicit relationships for the cost minimizing values of modal split and bus service (e.g., frequency), and a fixed number of commuters is assumed. The performed numerical simulations show that reserved bus lanes would not substantially contribute to the total cost reduction if marginal cost bus fares and car tolls could be charged. In Small (1983a), a disaggregate logit model for work trips is considered (also with a fixed number of commuters), and some of the service variables associated with bus (fare, headway and occupancy) are assumed to be constant. The numerical results in this case show that, even if less effective than congestion tolls, priority measures for buses deserve serious consideration as a strategy to deal with urban highway congestion.

Small (1983b) provides one of the first studies tackling the issue of incidence of congestion tolls (or equity). According to the author, such incidence cannot be logically analyzed without an explicit assumption about the use of revenues. The model assumes exogenous PT (bus) prices and service levels. A previously estimated logit model of mode choice for work trips in the San Francisco Bay Area is employed here, which allows to examine distributional effects. The obtained results show that in the absence of any redistribution of toll revenues, the welfare effects fall with income. Kraus (1989) also evaluates the welfare gains by studying different regimes that correspond to road pricing schemes that are technologically feasible: gas tax only, tolls via automatic vehicle identification and tolls via time-based meters installed on cars. The demand is modeled via households, whose utility function is Cobb-Douglas. The policy variables of the simulation model include bus fare and headway, as well as variables specific to the regime.

In addition to congestion, Borger and Wouters (1998) take into account all relevant external effects (air pollution, noise and accident risks) in a model that jointly optimizes transportation prices (for car and buses in peak and off-peak periods) and service decisions (bus supply of vehicle-kilometers in both periods, fleet size and capacity of the road infrastructure). The behavior of households is represented with deterministic utility functions that account for the above mentioned external effects. The objective is to maximize a measure of social welfare defined over the household utilities plus government net revenues, evaluated at the marginal cost of public funds. The model is tested with simulation experiments that consider aggregated Belgian data, which does not allow for distributional considerations. The results show that when the budgetary constraints are ignored, optimal prices were found well above the benchmark level for car and much lower for bus, and optimal supply of bus was substantially above. As expected, in the presence of budget constraints, prices in all modes and periods rise with respect to the benchmark model.

More recently, Basso et al. (2011) analyze urban congestion management policies (congestion pricing, transit subsidies and dedicated bus lanes) with a simple model that allows travelers to choose between car, bus and biking. The users' heterogeneity is addressed with a simple framework with the objective of increasing tractability that assumes that all users share the same value of time and income but differ in their valuation of some other attributes such as safety, comfort, etc. Such difference is included as the product of a modal-specific constant capturing these attributes and an idiosyncratic term that varies across the population (uniformly distributed random parameter). The authors evaluate different scenarios based on the considered policies by adding or ignoring restrictions on revenue recycling, namely self-financing bus system and the financing of bus fares from congestion tolls. The optimization problems associated with each scenario are numerically solved for a parameter setting representing a morning-peak hour in Santiago, Chile.

Note that demand data was not available and had to be artificially generated such that the modal split was reasonable. By comparing the resulting service levels, social welfare and consumer surplus, dedicated bus lanes provide a better stand-alone policy than PT or congestion pricing and it would count with public support.

Similarly, Basso and Silva (2014) assess the efficiency and substitutability of the same urban transportation policies, with a special emphasis on PT subsidies. In this case, the available demand data with respect to 5 different income groups in the population allows to perform distributional analyses with respect to the consumer surplus. The main results suggest that there is large efficiency substitutability between policies. This case study is further explained in Section 4.5, where a linearization scheme of the optimization problem associated with one of the evaluated policies is formulated and tested.

In summary, we have identified various works in the literature that simultaneously optimize the pricing and service design of a transportation system. Nevertheless, despite the importance of modal split effects, we observe relevant simplifications when it comes to represent the demand. This is related to the fact that the associated consumer surplus formulation, and the measure of social welfare derived thereof, becomes complex. As described in Section 4.3, the choice-based optimization framework provides a flexible framework for the linear formulation of the consumer surplus, as well as revenue recycling mechanisms. Furthermore, the disaggregate nature of the demand representation enables the development of distributional analyses with respect to differentiated segments in the population. This provides a tool that allows to model different contexts in order to evaluate alternative policies with respect to the public authority's established objectives.

4.3 Methodology

In this section, we rely on the choice-based optimization framework to model the problem of the welfare-maximizing pricing and service design of a transportation system. Thanks to the demand representation described in Section 2.2, we are able to incorporate not only the heterogeneous behavior of the travelers, but also the impact of the decision variables on the modal split and vice versa. The resulting formulation provides a tool to evaluate the implications on the welfare of different transportation policies. Moreover, the disaggregate nature of the demand allows to analyze distributional impacts among different groups in the population with similar socioeconomic characteristics.

We consider a population of N individuals that aim at performing a trip within a certain time horizon H (expressed in hours). The choice set $\mathcal{C}$ comprises, at least, two alternatives: PMM and PT. Both modes are managed by a single operator, such as a transportation authority or the government. We refer the reader to Bortolomiol et al. (2019) for the modeling framework that involves multiple operators. We assume that both alternatives are offered to all individuals and do not interact with each other, i.e., we do not account for mixed traffic conditions. The choice-based optimization framework allows to deal with capacity of the services (e.g., number of seats in a train), either given or as decision variables (see Section 2.2.3). As capacity introduces a high complexity to the formulation, and the nature of the case studies considered in this chapter do not explicitly involve capacity restrictions, we assume unlimited capacity in the case studies described in Sections 4.4 and 4.5.

We note that the travel time of a transportation mode is typically affected by congestion effects, especially in the case of PMM. Thus, travel time is a quantity that depends on the demand. Congestion is out of the scope for the choice-based optimization framework, as it is highly non-linear and it involves the explicit modeling of equilibrium. In this chapter, we provide simple modeling solutions to accommodate congestion effects in the case studies developed in Sections 4.4 and 4.5. We leave more advanced representations of congestion within the choice-based optimization framework for further research.

We might also allow for an opt-out option (denoted by $i = 0$), i.e., an alternative that captures the individuals that decided not to travel or that travel using a transportation mode that is not managed by the operator, such as slow modes (bike or walk) or modes managed by a different operator. Notice that it is possible to take into account differentiated PMM (e.g., car, motorcycle) and/or PT modes (e.g., train, bus, metro) as long as they are handled by the same operator.

The behavior of travelers is modeled with the mixed-integer linear representation of a DCM introduced in Section 2.2.2. Recall that the utility associated by individual n with alternative $i \in \mathcal{C}_n$, U_{in}, is composed of a systematic component V_{in} and an error component ε_{in}. We generate R draws from the probability distribution of the random component, ξ_{inr}, which allows to specify a deterministic utility associated with each draw r, individual n and alternative $i \in \mathcal{C}_n$ as $U_{inr} = V_{in} + \xi_{inr}$. The preference structure of travelers is characterized by a set of linear constraints on such utilities thanks to the choice variables w_{inr}. When the capacity of the alternatives is not considered and all alternatives are offered to all travelers by the operator (as is the case in Sections 4.4 and 4.5), the formulation can be simplified. We can disregard the discounted utility variables z_{inr} and determine the choice of each individual and each scenario with the variables representing the largest utility, U_{nr}, defined as $U_{nr} = \max_{i \in \mathcal{C}_n} U_{inr}$.

The authority needs to set the prices associated with both alternatives, i.e., a toll for PMM and a fare for PT. As in the previous chapters, we denote the price variables by p_{in}, where i refers to the alternative (PMM or PT) and n to the individual (see Section 2.3.1 for the modeling of the price in the continuous case). Recall that this price representation also embraces the definition of different prices for different groups of travelers or at the extreme, even a single price for everyone. Another variable to be optimized by the operator is the frequency of PT, denoted by f, which is assumed to be discrete, and represents the number of PT units (e.g., bus, trains) that are operated per hour. We define L_f frequency levels, denoted by f_ℓ, which represent feasible values for the frequency (as determined by the public authority and/or other considerations), and binary variables τ_ℓ, which are equal to 1 if frequency level f_ℓ is selected, and 0 otherwise, and define f as

$$f = \sum_{\ell=1}^{L_f} f_\ell \tau_\ell. \tag{4.1}$$

We need to impose that only one frequency level is selected, i.e.,

$$\sum_{\ell=1}^{L_f} \tau_\ell = 1. \tag{4.2}$$

These decisions are endogenous variables of the formulation, that is, they are present in the utility functions, and therefore have an impact on the behavior of the individuals. Note that additional design variables, such as decisions on the capacity or the scheduling of PT, can also be accommodated in the model.

The objective function to be maximized by the authority is a measure of social welfare, which is described in Section 4.3.1. The budgetary constraints that allow for a revenue recycling mechanism are depicted in Section 4.3.2. In Section 4.3.3, we provide a brief overview of the different strategies that allow for a linear characterization of the demand-dependent travel time. We summarize the proposed model-based algorithmic approach in Section 4.3.4.

4.3.1 Social welfare

As reviewed in Section 4.2, the welfare measures considered to evaluate the performance of a policy revolve around the consumer surplus. In the microeconomic consumer theory, consumer surplus is defined as the difference between what a consumer is willing to pay for one good and what they actually pay for it, which corresponds to the experienced satisfaction.

The difference in consumer surplus measures the changes in the values of the utility functions induced by alterations in the choice set and/or the attributes of the alternatives. Such difference can be calculated for DCM that are translationally invariant, which means that a shift in the systematic component of the utility function simply translates the joint distribution of the utilities without altering its basic functional form. Most of the DCM used in practice are translationally invariant or can be transformed into an equivalent translationally invariant model. In the following, we assume DCM with this property.

As noted by Williams (1977), the difference in the consumer surplus of an individual between two situations can be expressed as the difference between the associated expected maximum utilities. The expected maximum utility represents the average benefit obtained by an individual from the available alternatives. This quantity is individual-specific, reflecting the differences across individuals on how they evaluate the alternatives. Hence, as it provides the expected *worth* of a set of alternatives, Ben-Akiva and Lerman (1985) refer to it as a measure of accessibility.

In the case of the logit model, the expected maximum utility is calculated as

$$\mathbf{E}\left[\max_{i\in\mathcal{C}_n} U_{in}\right] = \frac{1}{\mu}\left(\ln\sum_{i\in\mathcal{C}_n} e^{\mu V_{in}} + \gamma\right), \qquad \forall n, \tag{4.3}$$

where μ is the scale parameter of the extreme value distribution (probability distribution of the error terms) and γ is the Euler's constant. This quantity is equivalent to the individual's consumer surplus up to a constant. The difference between the expected maximum utilities of two situations (denoted by 1 and 2) is equal to the difference in consumer surplus as the constant term cancels out:

$$\frac{1}{\mu}\left(\ln\sum_{i\in\mathcal{C}_n} e^{\mu V_{in}^2} - \ln\sum_{i\in\mathcal{C}_n} e^{\mu V_{in}^1}\right), \qquad \forall n, \tag{4.4}$$

where V_{in}^1 and V_{in}^2 represent the values of the utility functions for situations 1 and 2, respectively.

More generally, multivariate extreme value (MEV) models represent a family of DCM that allow for correlation among the error terms of the utility functions (McFadden, 1978). The logit model, the nested logit model and the cross-nested logit model are examples of MEV models. They are derived from the assumption that the error terms $\varepsilon_n = (\varepsilon_{1n}, \ldots, \varepsilon_{J_n n}), \forall n, J_n = |\mathcal{C}_n|$, follow a multivariate extreme value distribution, which corresponds to the multivariate version of the extreme value distribution leading to the logit model. The expected maximum utility of MEV models is computed by replacing the sum in (4.3) by the corresponding choice probability generating function (CPGF), i.e.,

$$\mathbf{E}\left[\max_{i\in\mathcal{C}_n} U_{in}\right] = \frac{1}{\mu}\ln G\left(e^{V_{1n}}, \ldots, e^{V_{J_n n}}\right), \qquad \forall n, \tag{4.5}$$

where $G\left(e^{V_{1n}}, \ldots, e^{V_{J_n n}}\right)$ denotes the CPGF. This function needs to satisfy some properties (the strong alternating sign property, the μ-homogeneity property and the limit property) in order to define a cumulative distribution function (CDF) of a multivariate extreme value distribution. Notice that if the G function has a closed form, so have the choice probabilities. For instance, the CPGF of the nested logit model, which is the

DCM of the case study considered in Section 4.5, is the following:

$$G\left(e^{V_{1n}}, \ldots, e^{V_{J_n n}}\right) = \sum_{m=1}^{M} \left(\sum_{j \in \mathcal{C}} (\alpha_{jm} e^{V_{jn}})^{\mu_m} \right)^{\frac{\mu}{\mu_m}}, \qquad \forall n, \tag{4.6}$$

where M denotes the number of nests in the model, α_{jm} is equal to 1 if alternative j belongs to nest m, and 0 otherwise, and μ and μ_m represent the scale parameter at the upper and lower level, respectively.

Although the expected maximum utility has a relatively tractable form for the logit model, it is non-linear. More advanced DCM, such as MEV models, present an even more complex formulation. We can linearly approximate the expected maximum utility of individual n thanks to the variables U_{nr} that are available in the choice-based optimization framework (see definition (2.22)). The simulation-based representation of DCM allows to approximate the expectation of $\max_{i \in \mathcal{C}_n} U_{in}$ with the average of U_{nr} across R, i.e.,

$$\mathbf{E}\left[\max_{i \in \mathcal{C}_n} U_{in}\right] \approx \frac{1}{R} \sum_{r=1}^{R} U_{nr}, \qquad \forall n. \tag{4.7}$$

We consider the aggregation across individuals as the measure of social welfare to be optimized:

$$\frac{1}{R} \sum_{n=1}^{N} \sum_{r=1}^{R} U_{nr}. \tag{4.8}$$

The nice property of this expression is that it is a linear function of the decision variables U_{nr}, so that it can be embedded in any MILP formulation. Notice that other measures of social welfare can be derived from (4.8), such as a the sum of the consumer surplus and the transportation authority surplus defined by the financial result of the transportation system (i.e., the difference between the obtained revenues and the experienced costs).

4.3.2 Budgetary constraint

In the absence of constraints limiting the investments to be made by the transportation authority, the optimization problem defined by the maximization of the consumer surplus (4.8) becomes meaningless, in the sense that the supply-related decisions (congestion toll, bus fare and frequency) will reach the associated bounds since the aggregated expected maximum utility increases as the investment on such decisions increases. Thus, we need to include a constraint ensuring that the investment does not exceed the available budget.

Notice that a budgetary constraint does not necessarily involve a revenue recycling mechanism, as the transportation authority might not consider the collected revenues as part of the budget dedicated to cover the performed investments. In this case, we assume that the revenues collected from the tolls and fares, together with an initial budget at hand (it can be set to 0 if not available), define the budget, and that the investment comprises various costs associated with the tolling infrastructure and the PT operation. Notice that the quantities defined below are in the same monetary units (e.g., CHF) and refer to the time horizon H under consideration.

The budget B is composed of an initial available budget B^0 and the collected revenues, i.e.,

$$B = B^0 + \frac{1}{R} \sum_{i \in \mathcal{C} \setminus \{0\}} \sum_{n=1}^{N} \sum_{r=1}^{R} \eta_{inr}, \tag{4.9}$$

where η_{inr} is the continuous variable that captures the product of variables $p_{in} w_{inr}$ (see constraints (2.39)–(2.42) in Section 2.3.1).

The investment I is composed of the investments I^{PMM} and I^{PT}. We assume that the former involves a fixed cost F^{PMM}, that represents the operating costs associated with the toll facility, and a variable cost c^{PMM}, that represents the transaction cost to be paid by the transportation authority for each individual that makes use of the toll facility. This cost is assumed to be the same for everyone. For the sake of simplicity, we only take into account a variable cost for PT, c^{PT}, associated with each operated PT unit (e.g., bus, train), which is calculated using the frequency f. The total investment I is calculated as

$$I = I^{\mathrm{PMM}} + I^{\mathrm{PT}} = F^{\mathrm{PMM}} + \frac{1}{R} c^{\mathrm{PMM}} \sum_{n=1}^{N} \sum_{r=1}^{R} w_{\mathrm{PMM},nr} + c^{\mathrm{PT}} H \sum_{\ell=1}^{L_f} f_\ell \tau_\ell. \tag{4.10}$$

The variable costs associated with PMM are calculated by multiplying the so-called transaction cost by the expected PMM demand. The variable cost associated with PT is obtained as the product of the cost associated with each PT unit by the total number of operated PT units, which is the product of the frequency (number of PT units per hour) and the length of the time horizon H, which is expressed in hours.

Thus, the budgetary constraint $I \leq B$ added to the model is the following:

$$F^{\mathrm{PMM}} + \frac{1}{R} c^{\mathrm{PMM}} \sum_{n=1}^{N} \sum_{r=1}^{R} w_{\mathrm{PMM},nr} + c^{\mathrm{PT}} H \sum_{\ell=1}^{L_f} f_\ell \tau_\ell \leq B^0 + \frac{1}{R} \sum_{i \in \mathcal{C} \setminus \{0\}} \sum_{n=1}^{N} \sum_{r=1}^{R} \eta_{inr}. \tag{4.11}$$

It basically assumes that the authorities are willing to invest a maximum of B^0 in addition to the revenues generated by the transportation system itself.

Notice that additional assumptions with respect to the budget and the investment can be included in the formulation as long as it remains linear. For instance, it is possible to impose self-financing of the PT system, i.e.,

$$c^{\mathrm{PT}} H \sum_{\ell=1}^{L_f} f_\ell \tau_\ell \leq \frac{1}{R} \sum_{n=1}^{N} \sum_{r=1}^{R} \eta_{\mathrm{PT},nr}, \tag{4.12}$$

or the financing of the PT fares with the road tolls, i.e.,

$$c^{\mathrm{PT}} H \sum_{\ell=1}^{L_f} f_\ell \tau_\ell \leq \frac{1}{R} \sum_{n=1}^{N} \sum_{r=1}^{R} \eta_{\mathrm{PMM},nr}. \tag{4.13}$$

4.3.3 Demand-dependent travel time

Since an increase in the demand of a transportation mode (especially in the case of PMM) typically induces an increase in its travel time, it is important to take into account congestion effects in the formulation. To this end, we rely on the standard function introduced by the US Bureau of Public Roads (Federal Highway, 1980), known as BPR function. For each arc, the travel time is provided by

$$tt(Y) = t_f \left(1 + a \left(\frac{Y}{C}\right)^b\right), \tag{4.14}$$

where t_f denotes the free-flow travel time (i.e., a fixed travel time when the arc is non congested), Y denotes the flow on the arc [veh/h] and C denotes its practical capacity [veh/h]. The practical capacity of an arc represents a threshold value for the flow beyond which congestion becomes effective. Notice that (4.14) is expressed in the same units as t_f (e.g., hours). The parameters a and b are set to the standard values suggested in Sheffi (1984), i.e., $a = 0.15$ and $b = 4$. Hence, when the flow Y is lower than C, the second term in parentheses in (4.14) becomes negligible, whereas for larger values of Y the non-linear contribution related to congestion increases. We note that the concept of practical capacity is different from the capacity restrictions associated with the choice-based optimization framework, as the practical capacity can be exceeded.

This formulation is usually employed to calculate the travel time by PMM. However, it can also be considered in the case of buses, as is done in the case study described in Section 4.5. Notice that the usage of the road by PMM involves the payment of a toll, whose collection might as well result in an increase of the travel time depending on the collection mechanism being employed. For the sake of illustration, we assume that such an increase is negligible, as is the case of wireless systems that automatically gather tolls (e.g., electronic toll collection).

The disaggregate demand representation allows to characterize the expected flow of transportation mode i (Y_i) via the choice variables w_{inr}. Notice that the choice variables are used to calculate the expected demand, and such demand allows to calculate the flow by translating the number of individuals into vehicles (via the vehicle occupancy) and by dividing by H (time horizon). As travel time is one of the variables explaining the choice of travelers, it is included in the utility function U_{inr}. Moreover, the utilities determine the choice variables. Thus, the demand depends on the travel time, and the travel time also depends on the demand, so there is a fixed-point problem.

As shown in Bortolomiol et al. (2019), because of the discrete and disaggregate nature of the choice-based optimization framework, an equilibrium of the problem might not exist. Hence, imposing equilibrium constraints would lead to an infeasible problem. It would be possible to consider a dynamic specification, where the decisions made at each draw are based on state variables generated at previous draws. The implication of this approach is that the resulting formulation is no longer draw-independent, and decomposition methods such as those described in Chapter 3 cannot be used anymore.

For all these reasons, we have decided to develop a fixed-point iterative algorithm, as is done in Bortolomiol et al. (2019). The idea is to solve a sequence of optimization problems where the flow Y_i is exogenous and determined by the choice of individuals as provided by the solution of the previous problem in the sequence. This iterative procedure is referred to as *Banach iterations*, and might or might not converge. In the absence of an equilibrium, it may even be a cycle. We use it as a heuristic approach to account for congestion effects in the case studies described in Sections 4.4 and 4.5.

The flow Y_i^0 is initialized to some value and the flow Y_i^k at iteration $k > 0$ is defined as

$$Y_i^k = \frac{D_i^{(k-1)}}{H\varrho} = \frac{1}{RH\varrho}\sum_{n=1}^{N}\sum_{r=1}^{R}\bar{w}_{inr}^{(k-1)}, \tag{4.15}$$

where $D_i^{(k-1)}$ represents the demand of mode i at iteration $k-1$, ϱ denotes the vehicle occupancy (passengers per vehicle) and $\bar{w}_{inr}^{(k-1)}$ are the values of the choice variables obtained when solving the welfare-maximization problem with $Y^{(k-1)}$. The algorithm terminates when one of the following criteria is satisfied: a maximum number of iterations K is achieved, the relative difference between two consecutive optimal objective functions is lower than a threshold δ, or a cycle (i.e., two optimal objective functions are alternatively generated) is obtained. The resulting optimization problem, together with the pseudocode of the algorithm, is detailed in Section 4.3.4.

4.3.4 Model-based algorithmic approach

The welfare maximization problem is formulated in Model 4.1. Constraints (4.17) determine the utility function associated with each individual n and draw r for each transportation mode i, constraints (4.18)–(4.21) provide the linear definition of the discounted utility variables, constraints (4.22)–(4.25) characterize the choice in terms of the discounted utility variables, constraints (4.26)–(4.27) handle capacity restrictions on the transportation modes, constraints (4.28)–(4.31) provide a linear formulation of the product of variables $p_{in}w_{inr}$, constraint (4.32) defines the frequency, and constraints (4.33)–(4.35) refer to the budgetary constraint.

$$
\begin{aligned}
SW(Y) = \max \quad & \frac{1}{R}\sum_{n=1}^{N}\sum_{r=1}^{R} U_{nr} && && (4.16)\\
\text{s.t.} \quad & U_{inr} = \beta^{p}_{in}p_{in} + \beta^{t}_{in}tt(Y_i) + \beta^{f}_{in}f + d_{inr} && \forall n, i \in \mathcal{C}_n, r && (4.17)\\
& \ell_{nr} \le z_{inr} && \forall i \in \mathcal{C}, n, r && (4.18)\\
& z_{inr} \le \ell_{nr} + M_{inr}y_{inr} && \forall i \in \mathcal{C}, n, r && (4.19)\\
& U_{inr} - M_{inr}(1 - y_{inr}) \le z_{inr} && \forall i \in \mathcal{C}, n, r && (4.20)\\
& z_{inr} \le U_{inr} && \forall i \in \mathcal{C}, n, r && (4.21)\\
& z_{inr} \le U_{nr} && \forall i \in \mathcal{C}, n, r && (4.22)\\
& U_{nr} \le z_{inr} + M_{inr}(1 - w_{inr}) && \forall i \in \mathcal{C}, n, r && (4.23)\\
& \sum_{i\in\mathcal{C}} w_{inr} = 1 && \forall n, r && (4.24)\\
& w_{inr} \le y_{inr} && \forall i \in \mathcal{C}, n, r && (4.25)\\
& \sum_{m=1}^{n} w_{imr} \le (c_i - 1)y_{inr} + (n-1)(1 - y_{inr}) && \forall i \in \mathcal{C}, n > c_i, r && (4.26)\\
& c_i(1 - y_{inr}) \le \sum_{m=1}^{n} w_{imr} && \forall i \in \mathcal{C}, n > 1, r && (4.27)\\
& a_{in}w_{inr} \le \eta_{inr} && \forall i \in \mathcal{C}\setminus\{0\}, n, r && (4.28)\\
& \eta_{inr} \le b_{in}w_{inr} && \forall i \in \mathcal{C}\setminus\{0\}, n, r && (4.29)\\
& p_{in} - (1 - w_{inr})b_{in} \le \eta_{inr} && \forall i \in \mathcal{C}\setminus\{0\}, n, r && (4.30)\\
& \eta_{inr} \le p_{in} - (1 - w_{inr})a_{in} && \forall i \in \mathcal{C}\setminus\{0\}, n, r && (4.31)\\
& f = \sum_{\ell=1}^{L_f} f_\ell \tau_\ell && && (4.32)\\
& B = B^0 + \frac{1}{R}\sum_{i\in\mathcal{C}\setminus\{0\}}\sum_{n=1}^{N}\sum_{r=1}^{R}\eta_{inr} && && (4.33)\\
& I = F^{\text{PMM}} + \frac{1}{R}c^{\text{PMM}}\sum_{n=1}^{N}\sum_{r=1}^{R} w_{\text{PMM},nr} + c^{\text{PT}}Hf && && (4.34)\\
& I \le B && && (4.35)\\
& w_{inr}, \tau_\ell \in \{0, 1\}, && \forall i \in \mathcal{C}, n, r, \ell && (4.36)
\end{aligned}
$$

Model 4.1: Welfare maximization problem

The pseudocode of the iterative algorithm is presented in Algorithm 4.1.

Algorithm 4.1: Fixed-point iterative algorithm for Model 4.1

```
Input: Number of iterations K, threshold δ;
Output: A list of fixed-point solutions;
Initialize Y^k for k = 0, solve Model 4.1 with Y = Y^0;
while k < K or |(SW(Y^{k-1}) - SW(Y^k)) / SW(Y^k)| < δ or {SW(Y^k), SW(Y^{k-1})} define a cycle
do
    Solve Model 4.1 with Y = Y^k and obtain the values of the choice variables
    (w̄^k_inr);
    Set k ← k + 1;
    Calculate Y^k using (4.15) with w̄^{k-1}_inr;
```

4.4 Introduction of a congestion toll in an urban context

In this section, we define a small but meaningful case study to test the methodology introduced in Section 4.3. Similarly to Section 2.4, we rely on a DCM that has been calibrated and validated by experts. More precisely, we consider the integrated choice and latent variable (ICLV) model estimated in Atasoy et al. (2013) on travel mode choice in suburban areas of Switzerland (an overview of the DCM is provided in Section 4.4.1). Such areas are typically served by PostBus, the PT branch of the Swiss postal service.

Within the framework of the PostBus project, a Revealed Preference (RP) survey was sent to households of several towns and villages considered to be representative of the PostBus network. The respondents were asked to report information about all trips performed in a day, together with opinions on topics related to environment, lifestyle, etc., and data about their mobility habits, household composition and other socioeconomic characteristics.

As the data gathered in the survey involves multiple trips per individual and regions across Switzerland, we define a particular setting consisting of a single origin region representing a residential area and a single destination region representing a workplace/study place and a one-hour time horizon. We designate the residential area in Morges and the workplace/study place in Lausanne, being both cities in the French-speaking canton of Vaud (Switzerland). These two cities are at the core of the *Projet d'agglomération Lausanne-Morges* (PALM), which is a project on the urban development of Lausanne and its suburbs and the city of Morges. The aim of this project revolves around the densification of the regional urban tissue at the same time that an excessive urban sprawl is avoided, being one of the main focus the improvement of the public transportation network.

In order to set a particular origin region within Morges and a particular destination region within Lausanne, we get inspiration from the geographical definition of regions in the PALM project for which the incoming, outgoing and inner trips are measured. Figure 4.1 shows the different regions (referred to as *cordons* in French). In our case, we consider trips with origin in the city center of Morges (region 4) and destination in the city center of Lausanne (region 1).

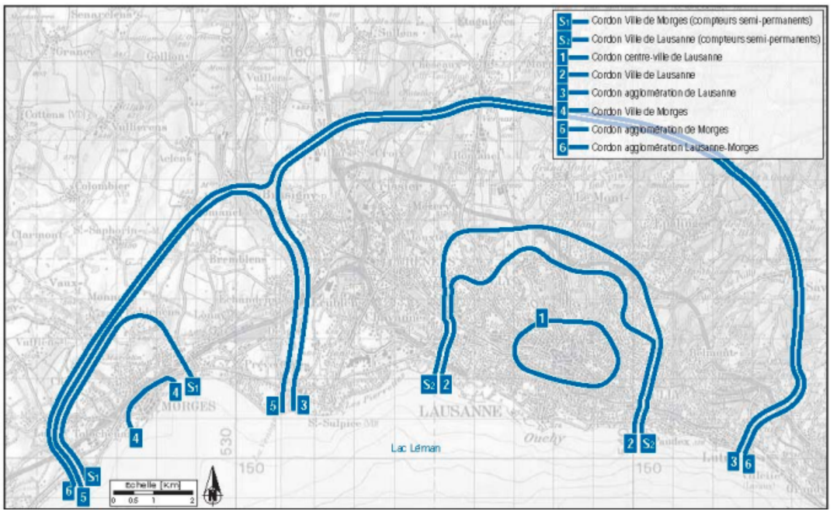

Figure 4.1: Geographical division of the Lausanne-Morges region into subregions. Source: Guillaume-Gentil et al. (2015)

The choice of these cities and the corresponding origin and destination regions is motivated by various reasons. First, we can assume that the alternative PMM involves the use of the highway (and therefore individuals can be charged with a toll), since approximately 70% of the trips by PMM within the PALM region (region 6 in Figure 4.1) use the highway (Guillaume-Gentil et al., 2015). Second, we can assume that the PT alternative only comprises the train, as it provides the fastest connection and is the only mode linking the two regions without the need for transfers (without taking into account a potential additional connection by PT in the destination region for some individuals in order to reach their final destination, see Section 4.4.2 for more details). Finally, and given the distance between both regions (between 11 and 15 km depending on the exact origin and destination of the trip), bicycle is assumed to be available to individuals, which enables the inclusion of the SM alternative as an available transportation mode.

In Section 4.4.2, we create a semi-synthetic population to test the optimization problem detailed in Section 4.3.4 on the defined setting. To this end, we rely on the Mobility and Transport Microcensus (MTMC) survey, which is a statistical survey of the travel behavior of the Swiss population that is conducted every five years by the Federal Office for Spatial Development (ARE) and the Federal Statistical Office (FSO). We select the respondents with residency in Morges, and we identify the questions of interest in the survey in order to determine the necessary socioeconomic variables of the ICLV model. As for the generation of trips, we perform some assumptions on the attributes that are present in the ICLV model based on the specifications of the setting.

Notice that the design variables (PMM toll and PT fare and frequency) are to be decided by the optimization problem. Section 4.4.3 establishes a base case that will be considered as a benchmark, and Section 4.4.4 discusses the results obtained from solving the welfare maximization problem (using Algorithm 4.1) for three different situations with respect to the level of congestion of the highway. All the experiments have been implemented in C++ using ILOG Concert Technology to access CPLEX 12.8, and all the instances were performed using 12 threads in a 3.33 GHz Intel Xeon X5680 server running a 64-bit Ubuntu 16.04.2.

4.4.1 ICLV model

The purpose of incorporating latent attitudes of the individuals into a travel mode choice model is to better understand their underlying choice preferences in order to increase the forecasting power of the DCM. As mentioned in Section 2.2.1, attitudinal variables are included into a choice context through latent variables, latent classes or a combination thereof. In ICLV models, the attitudinal variables are included as explanatory variables in the DCM, whereas integrated choice and latent class (ICLC) models use the attitudinal variables to identify groups that might have different taste parameters, choice sets, or decision protocols. These latent variables and classes are identified with the so-called psychometric indicators, which are statements on opinions for which a level of agreement needs to be evaluated by the respondents.

In both cases, structural equation models allow to translate attitudinal variables into a statistical model. Such models are integrated into DCM in order to implement a simultaneous estimation of the choice and attitudinal variables. Hence, as the heterogeneity in the sample is explained through structural equation models for attitudinal variables, the resulting models can be applied for other samples provided that the required variables are available.

In Atasoy et al. (2013), an ICLV model and an ICLC model are calibrated and tested against a logit model that has the same model specification as the DCM comprised in both approaches. The obtained results show that the ICLC model has the best fit compared to the other two models. Nevertheless, as the ICLV model involves the calculation of integrals within the associated choice probabilities that do not have a closed-form (Walker, 2001), it is more challenging to derive welfare measures for this model. This is why we have decided to select the ICLV model to illustrate the extent of the proposed methodology.

Mode choice is assumed to be between PMM, which encompass car as a user and passenger, motorbike and taxi, PT, which consists of bus, train and PostBus, and SM, which represent walking or biking. The model considers two attitudes, named pro-car (denoted by $\mathrm{A_{car}}$) and environmental concern (denoted by $\mathrm{A_{env}}$). Tables 4.1 and 4.2 provide the specification table of the utilities and the structural equations for the attitudes, respectively. The index n that represents the individual has been omitted in the explanatory variables in order to simplify the notations. Recall that the specification table includes both the estimates of the parameters as well as the specification that define the equations (utilities and structural equations for the attitudes). $\mathrm{ASC_{PMM}}$ and $\mathrm{ASC_{SM}}$ are the constants for the utilities of PMM and SM ($\mathrm{ASC_{PT}}$ is set to 0 for normalization purposes), $\bar{\mathrm{A}}_{\mathrm{car}}$ and $\bar{\mathrm{A}}_{\mathrm{env}}$ are the constants for the corresponding attitudes, the parameters associated with the modal attributes and individual characteristics are denoted by β, and the parameters associated with the explanatory variables of the attitudes are denoted by ζ.

Table 4.1: Specification table of the utilities of the ICLV model (Atasoy et al., 2013)

		PMM	PT	SM
$\mathrm{ASC_{PMM}}$	-1.5	1	0	0
$\mathrm{ASC_{SM}}$	0.5	0	0	1
β_{cost}	-0.056	$\mathrm{C_{PMM}}$	$\mathrm{C_{PT}}$	0
$\beta_{\mathrm{TT_{PMM}}}$	-0.029	$\mathrm{TT_{PMM}}$	0	0
$\beta_{\mathrm{TT_{PT}}}$	-0.012	0	$\mathrm{TT_{PT}}$	0
$\beta_{\mathrm{distance}}$	-0.224	0	0	$\mathrm{D_{SM}}$
$\beta_{\mathrm{N_{cars}}}$	0.97	$\mathrm{N_{cars}}$	0	0
$\beta_{\mathrm{N_{children}}}$	0.215	$\mathrm{N_{children}}$	0	0
$\beta_{\mathrm{language}}$	1.06	French	0	0
β_{work}	-0.583	WorkTrip	0	0
β_{urban}	0.283	0	Urban	0
β_{student}	3.26	0	Student	0
$\beta_{\mathrm{A_{car}}}$	-0.574	0	$\mathrm{A_{car}}$	0
$\beta_{\mathrm{A_{env}}}$	0.393	0	$\mathrm{A_{env}}$	0
$\beta_{\mathrm{N_{bikes}}}$	0.385	0	0	$\mathrm{N_{bikes}}$

The explanatory variables included in the equations in Tables 4.1 and 4.2 are the following:

- travel cost ($\mathrm{C_{PMM}}$ and $\mathrm{C_{PT}}$) [CHF],
- travel time ($\mathrm{TT_{PMM}}$ and $\mathrm{TT_{PT}}$) [min],
- distance ($\mathrm{D_{SM}}$) [km],
- number of cars in the household ($\mathrm{N_{cars}}$),
- number of children in the household ($\mathrm{N_{children}}$),
- number of bikes in the household ($\mathrm{N_{bikes}}$),

Table 4.2: Specification table of the structural equations for the attitudes of the ICLV model (Atasoy et al., 2013)

		$\mathrm{A_{car}}$	$\mathrm{A_{env}}$
$\bar{\mathrm{A}}_{\mathrm{car}}$	3.02	1	0
$\bar{\mathrm{A}}_{\mathrm{env}}$	3.23	0	1
$\zeta_{\mathrm{N_{cars}}}$	0.104	$\mathrm{N_{cars}}$	0
ζ_{educ}	0.235	-Educ	Educ
$\zeta_{\mathrm{N_{bikes}}}$	0.085	0	$\mathrm{N_{bikes}}$
ζ_{age}	0.004	0	$\mathrm{Age} \cdot (\mathrm{Age} > 45)$
ζ_{Valais}	-0.223	Valais	0
ζ_{Bern}	-0.361	Bern	0
$\zeta_{\mathrm{Basel\text{-}Zurich}}$	-0.256	Basel-Zurich	0
ζ_{East}	-0.228	EastSwitzerland	0
$\zeta_{\mathrm{Graubünden}}$	-0.303	Graubünden	0

- an indicator that is equal to 1 if the respondent is French-speaking (French),
- an indicator that is equal to 1 if the purpose of the trip is going to work (Worktrip),
- an indicator that is equal to 1 if the residency of the individual is in an urban area (Urban),
- an indicator that is equal to 1 if the respondent is a student (Student),
- a set of indicators associated with different regions in Switzerland (Valais, Bern, Basel-Zurich, EastSwitzerland, Graubünden) that are equal to 1 if the residency of the individual is within the region,
- an indicator that is equal to 1 if the respondent has an university degree (Educ), and
- a piecewise linear variable that is equal to the individual's age if the age is larger than 45 years, and 0 otherwise ($\mathrm{Age} \cdot (\mathrm{Age} > 45)$).

Notice that the frequency of PT is not explicitly included in the ICLV model. In order to capture the interaction of this design variable with the preferences of the travelers, we express $\mathrm{TT}_{\mathrm{PT},n}$ as the sum of its main components, namely access time, waiting time, in-vehicle time and egress time. We formulate the waiting time in terms of the headway, which is the inverse of the frequency. A more detailed explanation is provided in Section 4.4.2. Hence, reducing or increasing the frequency will have an impact on the travel time $\mathrm{TT}_{\mathrm{PT},n}$, which will be captured by the ICLV model by means of the term $\beta_{\mathrm{TT_{PT}}}\mathrm{TT}_{\mathrm{PT},n}$ that is included in the utility function associated with PT. Note that this does not modify the ICLV model, as it simply affects the way $\mathrm{TT}_{\mathrm{PT},n}$ is calculated.

4.4.2 Assumptions about the demand

In this section, we design a semi-synthetic population by characterizing the variables that are present in the ICLV model. We rely on the MTMC data to obtain the individual characteristics and we generate the modal attributes in accordance with the defined setting. In the MTMC data, we identify $N = 100$ individuals that live in Morges. For the sake of simplicity, and in order not to check the actual location of the address of each respondent within the city, we assume that all individuals live in region 4.

As the highway is also being used by travelers with different origins and destinations, we need to adjust the flow Y_{PMM} that is obtained from the considered population to capture the background traffic. One way to do it is by associating a constant weight with each individual (assumed to be the same across individuals) in order to scale up the demand and represent the background conditions. The weights can be fine-tuned in order to represent different situations with respect to the existing level of congestion on the highway, which we define as non-congested, partially congested and heavily congested. The weights are determined based on the assumption associated with the practical capacity of the highway for the BPR function (4.14), which is described in Section 4.4.3.

Some of the necessary variables can be directly found on the MTMC data, whereas for some others we need to process the available information and perform some additional assumptions. The information on the purpose of the trip cannot be directly inferred as multiple trips are recorded per individual, so we assume that WorkTrip_n is equal to 1 if individual n is not a student, and 0 otherwise. The Student_n indicator is set to 1 in two cases: if the age of respondent n is less or equal than 16 or if they answered to be involved in an educational institution as a student.

Notice that the variables French_n and Urban_n are equal to 1 for all the individuals in the sample due to the characteristics of the scenario. Furthermore, all the region indicators (Valais_n, Bern_n, Basel-Zurich_n, EastSwitzerland_n, Graubünden_n) are set to 0 because Morges is in the canton of Vaud.

In order to generate the trip associated with each individual, we assume that the distance between their origin and their destination consists of a distance to be traveled within the origin zone (region 4) denoted by d_n^{M}, a distance to be traveled within the destination zone (region 1) denoted by d_n^{L}, and a distance between the two zones denoted by $\text{d}_n^{\text{M,L}}$. Hence, $\text{d}_{\text{SM},n} = \text{d}_n^{\text{M}} + \text{d}_n^{\text{M,L}} + \text{d}_n^{\text{L}}$.

Based on the origin and destination regions in Figure 4.1, we assume that d_n^{M} ranges between 0.1 and 1.5 km, whereas d_n^{L} varies between 0.2 and 3 km. For each individual in the sample, we impute both distances by drawing from a uniform distribution with the specified values as minimum and maximum. We fix the distance connecting the two zones to 12 km, i.e., $\text{d}_n^{\text{M,L}} = 12, \forall n$.

The travel time by PMM is obtained by calculating the travel time associated with each of the above defined distances. In the case of $\mathrm{d}_n^{\mathrm{M}}$ and $\mathrm{d}_n^{\mathrm{L}}$, we assume that the speed in an urban context is equal to 15 km/h (Christidis and Ibáñez-Rivas, 2012). The travel time corresponding to $\mathrm{d}_n^{\mathrm{M,L}}$ depends on the network conditions, and is determined with the BPR function via the iterative procedure described in Section 4.3.3. Thus, the travel time $\mathrm{TT}_{\mathrm{PMM},n}$ of each individual n is calculated by adding the three travel times.

As mentioned in Section 4.4.1, $\mathrm{TT}_{\mathrm{PT},n}$ is calculated as the sum of the access time from the individual's home to the train station in Morges, the waiting time of the individual at the train station, the in-vehicle time and the egress time from the Lausanne train station to the final destination of the individual. We define the access and egress time as the times spent to cover $\mathrm{d}_n^{\mathrm{M}}$ and $\mathrm{d}_n^{\mathrm{L}}$, respectively. For the former, we assume a walking speed of 5 km/h (i.e., common assumption of 1.4 m/s). For the latter, and given the size of region 1, we assume the 5 km/h walking-speed if the distance is shorter than 1.5 km, and the 15 km/h driving-speed in a urban environment if the distance is larger, representing the fact that the individual is not walking from the train station to their final destination but taking a faster mode (e.g., bus) instead. As the ICLV model does not take into account the departure time of the individual, we simplify the definition of in-vehicle time by using the same value for everyone: 13.8 minutes, which is the weighted average of the in-vehicle time of the trains connecting both cities during a one-hour time horizon in the current schedule (2 connections with a duration of 12 minutes, 2 with a duration of 13 minutes, 1 with a duration of 15 minutes and 1 with a duration of 18 minutes). Note that the travel times are different due to the intermediate stops of each connection.

Since the train timetables are available on multiple online platforms and are highly reliable in Switzerland, we can assume that individuals know the departure time of the train they aim at taking, and do not plan to arrive to the train station way in advance. Therefore, the usual assumption that defines the expected waiting time as half of the headway is not appropriate here. Instead, we define the waiting time $\mathrm{WT}_{\mathrm{PT},n}$ of individual n as (Ingvardson et al., 2018)

$$\mathrm{WT}_{\mathrm{PT},n} = \nu_n \cdot h, \tag{4.37}$$

where ν_n is drawn from a truncated Beta distribution with shape parameters equal to 3 and 8 (see Figure 4.2) and h denotes the headway. The distribution is truncated in the sense that ν_n can only take values between 0 and 1, since the waiting time should not exceed the headway. The headway is calculated as the inverse of the frequency multiplied by 60 in order to be expressed in minutes. Hence, applying (4.1), h is obtained as

$$h = \sum_{l=1}^{L_f} \frac{60}{f_l} \tau_l. \tag{4.38}$$

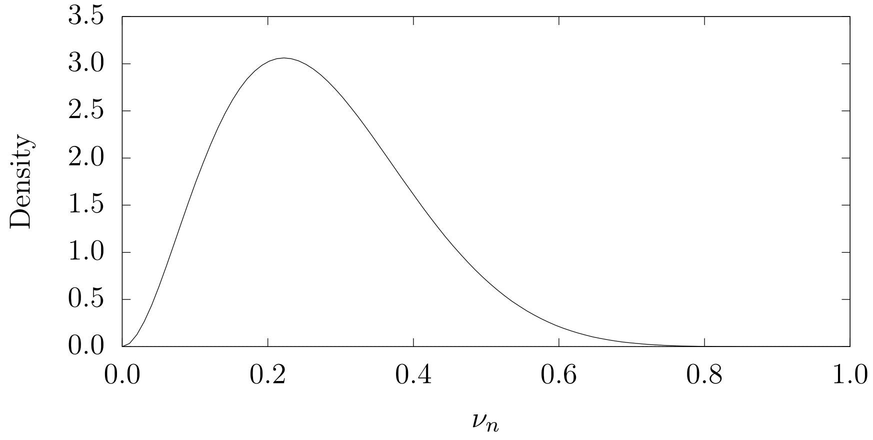

Figure 4.2: Distribution of ν_n in the definition of the waiting time (4.37)

In the next section, we specify the characteristics of the setting with respect to the practical capacity of the highway (for the travel time formulation (4.14)) and the costs associated with the PMM and PT alternatives. Furthermore, we use as a benchmark a base case that sets the design variables to their current values in reality.

4.4.3 Assumptions about the supply

In order to test the methodology introduced in Section 4.3, we make some assumptions on the practical capacity of the highway and the costs for the budgetary constraint (see Section 4.3.2). We rely on the general literature to determine such values to provide a realistic representation of the setting under consideration. We also define the base case used as a benchmark by fixing the design variables to the values that they currently have in reality. Notice that we do not consider the capacity constraints of the choice-based optimization framework, as we assume that there is room for all individuals in all transportation modes. The increase in demand in the case of PMM has an effect on the behavior of individuals because it affects the travel time via congestion effects. In the case of PT, however, the ICLV model does not account for sensitivity towards crowds, so such an effect is not captured.

In this case study, we assume that the highway used by PMM has 3 lanes, since the two highway stretches defining the route between Morges and Lausanne (A1 and E23) have 3 lanes each. According to TRB (2010), a 3-lane highway with a free-flow speed of 120 km/h has a practical capacity $C = 7200$ veh/h.

In the non-congested case, the flow Y_{PMM} does not exceed the practical capacity C, whereas in the partially and highly congested cases we assume an excess of 20% and 50%, respectively. With respect to the weights capturing the background traffic, they are derived for an approximated sample flow of 40 veh/h in the non-congested case and

30 veh/h in the congested cases, which means that the weights are equal to 180 for the non-congested situation, 288 for the partially congested case and 400 for the heavily congested case.

We need to set values for the costs defining the investment (4.10) in the budgetary constraint. The transaction cost c^{PMM} is set to CHF 0.44 (KPMG, 2015). The cost of operating each train is obtained from the range of marginal infrastructure cost in train-km used for policy purposes in MOVE (2014), which establishes such cost between CHF 0.75 and CHF 1.40 per train-km for passenger trains. As the distance to be covered is 12 km, we assume c^{PT} = CHF 10 (which takes a low value within the range such that the order of magnitude is appropriate to the sample size). As the fixed cost F_{PMM} might depend on the collection mechanism, we rely on an approximate cost per vehicle of CHF 1.25 (Ardekani, 1991), which allows us to define $F_{\mathrm{PMM}} = 50$ by considering the approximated sample flow of 40 veh/h.

In addition to the toll, PMM users need to pay for the variable costs associated with their vehicles, which include maintenance and repairs, tires, gas and depreciation. According to Touring Club Suisse (TCS), this cost corresponds to CHF/km 0.27.

For the base case used as a benchmark, we define the PT fare by dividing the cost of the monthly ticket that enables unlimited trips between Morges and Lausanne (CHF 137 for 2019) by the number of working (studying) trips in a month (252 working days in 2019 in canton Vaud, which makes an average of 21 working days per month, i.e., 42 trips per month), CHF 3.27[1], which is assumed to be the same for everyone. We assume that there is no congestion toll in this case, as the current road pricing scheme, the vignette, has a cost of CHF 40 (in 2019) for the whole year, so the cost per trip becomes negligible. With respect to the PT frequency, as mentioned in Section 4.4.2, we allow for a frequency of 6 trains per hour between both cities.

4.4.4 Numerical results

As is done for the base case, we assume that both prices are the same across individuals. Hence, we denote the highway toll simply by p_{PMM} and the PT fare by p_{PT}, and $p_i \in [0, 4]$ for $i \in \{\mathrm{PMM}, \mathrm{PT}\}$. The frequency levels are defined near the base case frequency: $f \in \{4, 5, 6, 7, 8\}$ (i.e., $L_f = 5$).

The social welfare associated with the base case is obtained by fixing $p_{\mathrm{PMM}} = 0$, $p_{\mathrm{PT}} = 3.27$ and $f = 6$ in the model-based algorithmic approach described in Section 4.3.4 for the three levels of congestion. We then run the approach without fixing the design variables and compare the obtained results. Similar to Section 2.4, we run Algorithm 4.1 on the

[1]This value is similar to the fare to be paid by travelers that own the Swiss half-fare ticket (CHF 3.70). The full fare (CHF 7.40) is rarely paid by commuters as they either have a monthly subscription or the half-fare ticket.

problem formulated in Model 4.1 for three replications (each replication corresponds to an independent generation of R draws of the error terms $\varepsilon_{in}, \forall i \in \mathcal{C}, n$).

Figures 4.3, 4.4, and 4.5 show the difference in social welfare between the optimized case and the base case in CHF in the non-congested, partially and highly congested situations, respectively. For the sake of illustration, if a cycle was reported by the iterative algorithm for one of the replications, we provide the average value of the social welfare of the two solutions. Notice that the social welfare can be expressed in monetary units by transforming the utility into monetary units, which can be done by dividing them by the cost coefficient β_{cost} of the ICLV model (see Table 4.2).

We can see that the difference in the social welfare tends to stabilize as the number of draws increases. Nevertheless, we still observe a certain degree of variability of the objective function for large values of R, which is expected due to the nature of the function being maximized (expected maximum utility). Moreover, as the level of congestion increases, the difference in social welfare experiences higher fluctuations. This is related to the way the travel time is calculated. The higher the level of congestion, the larger the weights associated with the individuals to capture the background traffic. Due to the nature of the BPR function (4.14), an additional car user in the non-congested situation has a much lower impact on the travel time than an additional car user in the highly congested case.

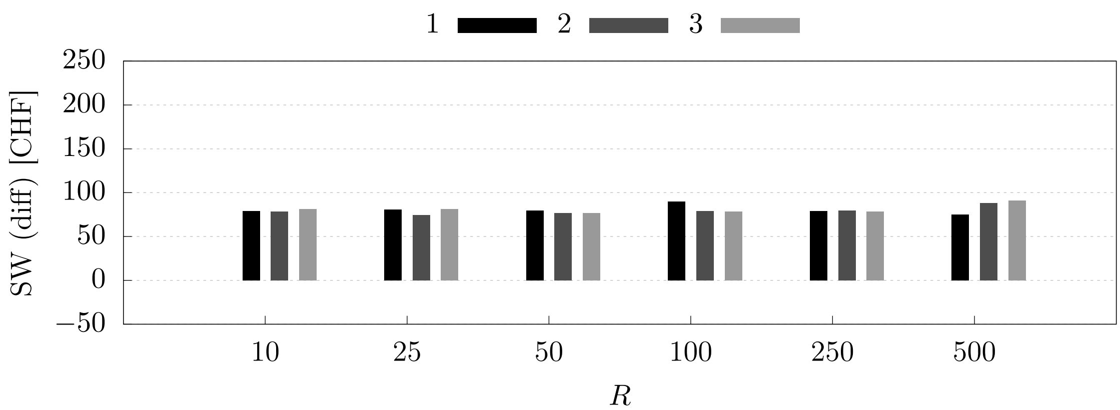

Figure 4.3: Difference in social welfare [CHF] with respect to the base case for the 3 considered replications (independent generation of draws, labeled as 1-3) for $R \in \{10, 25, 50, 100, 250, 500\}$ in the non-congested situation

As for the computational time, Table 4.3 shows the average running time (for the three considered replications) of Algorithm 4.1 for this case study (only the optimized case, as in the base case the design variables are fixed to the given values). We observe that the computational times for the partially congested case tend to be lower than in the other two cases, except for $R = 500$, which seems to be related to the number of iterations run by the algorithm. Despite the fact that an optimization problem needs to be solved at

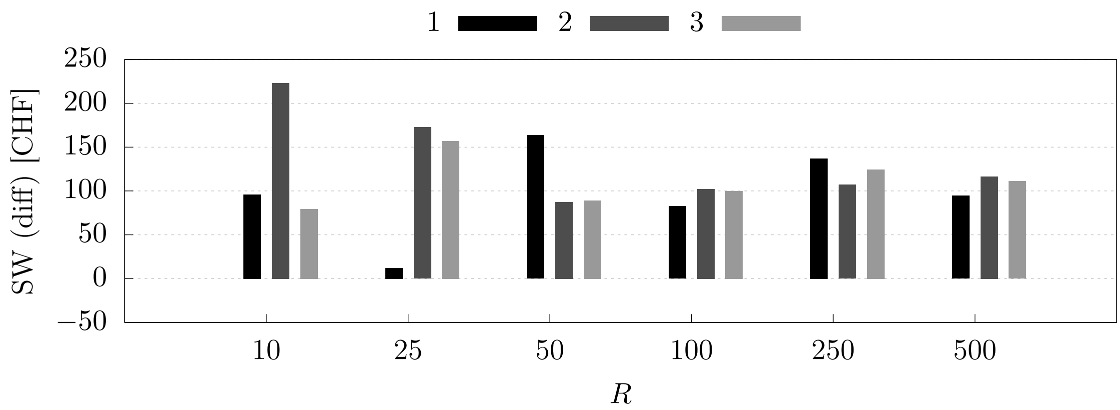

Figure 4.4: Difference in social welfare [CHF] with respect to the base case for the 3 considered replications (independent generation of draws, labeled as 1-3) for $R \in \{10, 25, 50, 100, 250, 500\}$ in the partially congested situation

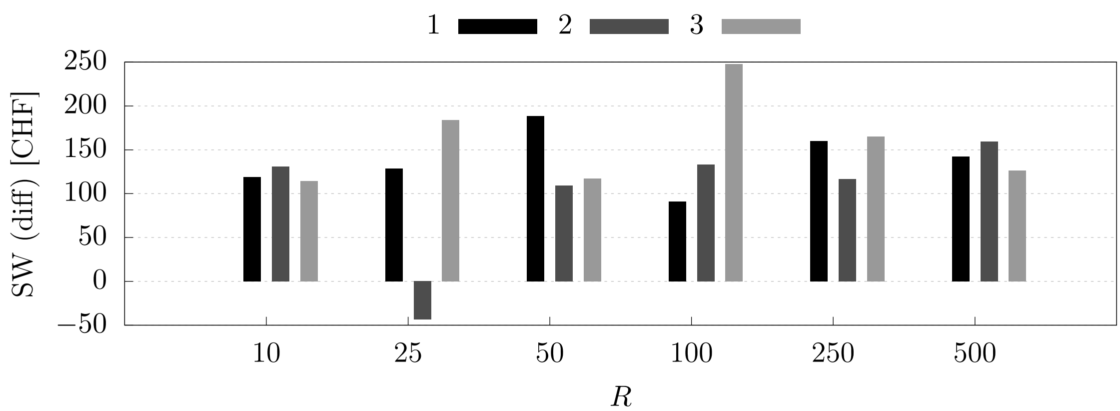

Figure 4.5: Difference in social welfare [CHF] with respect to the base case for the 3 considered replications (independent generation of draws, labeled as 1-3) for $R \in \{10, 25, 50, 100, 250, 500\}$ in the highly congested situation

each iteration, the solution times are not excessively high. This is due to the fact that capacity constraints are not considered in the formulation, which have a high impact on the computational complexity, as shown in Section 2.4.

Table 4.4 shows the values of the design variables determined by the algorithmic approach and Table 4.5 includes the modal split and the travel time for PMM for the base case and optimized case for the three replications and $R = 500$. More than one line per replication means that a cycle is reported by the algorithm (i.e., two solutions are generated).

We observe that the three design variables (p_{PMM}, p_{PT} and f) remain around the same values across replications and for the three levels of congestion. In particular, the frequency is always set to 5. Additional tests show that the increase of the cost associated with an additional train, c_{PT}, results in a lower frequency, which indicates that the model

Table 4.3: Average computational time (min) of Algorithm 4.1 for 3 replications (independent generation of draws) for the optimized case

R	Non-congested	Partially congested	Highly congested
10	0.118	0.0764	0.0781
25	0.580	0.417	0.706
50	2.86	3.60	4.82
100	20.5	15.0	38.1
250	147.0	99.3	182.0
500	351	1054	474

Table 4.4: Design variables (p_{PMM}, p_{PT} and f) for the 3 considered replications (independent generation of draws, labeled as 1-3) for $R = 500$ (multiple lines per replication indicate the presence of a cycle, i.e., the algorithm terminated because of a cycle and two solutions were reported)

Level of congestion		p_{PMM}	p_{PT}	f
Non-congested	1	1.50	1.06	5
Non-congested	2	1.43	1.10	5
Non-congested	3	1.50	1.06	5
Partially congested	1	1.50	1.06	5
Partially congested	1	1.44	1.10	5
Partially congested	2	1.49	1.07	5
Partially congested	3	1.50	1.06	5
Highly congested	1	1.51	1.07	5
Highly congested	2	1.42	1.10	5
Highly congested	3	1.44	1.09	5

is sensitive to cost. We notice that by fixing the frequency to the same value as in the base case, i.e., $f = 6$, the obtained congestion tolls and PT fares are above the values included in Table 4.4 (average values of CHF 1.60 for the congestion toll and CHF 1.18 for the PT fare), as a consequence of the additional train that is being operated, but the social welfare is slightly lower for all levels of congestion. Moreover, we note that having similar prices shows the specificity of this case (e.g., same price sensitivity across travelers and modes in the ICLV model, assumptions on the infrastructure and operating costs), as in general prices are used by the model to control the travel times.

The introduction of the congestion toll induces a decrease in the share of PMM with respect to the base case for the three levels of congestion. Such decrease is more noticeable in the non-congested situation, as the payment of the congestion toll has a low impact on the travel time savings, whereas this saving becomes larger in the other two situations. Similarly, the share of SM slightly decreases, as some users that were not willing to pay the PT fare in the base case might be willing to pay the current fare. Consequently, the share of PT increases as it captures the travelers coming from the other two modes.

Table 4.5: Modal split (%) and travel time for PMM [min] for the base case and the optimized case for the 3 considered replications (independent generation of draws, labeled as 1-3) for $R = 500$ (multiple lines per instance indicate the presence of a cycle, i.e., two solutions were reported)

Congestion level	Replication	D_{PMM}		D_{PT}		D_{SM}		$tt(Y_{\mathrm{PMM}})$	
		Base	Opt.	Base	Opt.	Base	Opt.	Base	Opt.
Non-cong.	1	38.4	33.6	55.7	60.5	5.85	5.88	12.4	11.8
Non-cong.	2	38.5	34.1	55.4	60.0	6.09	5.85	12.4	11.9
Non-cong.	3	38.4	34.0	55.7	60.3	5.93	5.73	12.4	11.9
Partially cong.	1	34.5	31.4	59.4	62.8	6.12	5.88	17.0	15.3
Partially cong.	1		31.7		62.4		5.87		15.1
Partially cong.	2	34.5	31.5	59.3	62.6	6.22	5.93	17.0	15.1
Partially cong.	3	34.5	31.4	59.3	62.8	6.26	5.88	17.0	15.2
Highly cong.	1	30.1	28.0	63.4	65.7	6.51	6.36	24.4	20.6
Highly cong.	1	30.4		63.1		6.48		24.0	
Highly cong.	2	30.1	28.3	63.2	65.4	6.67	6.30	24.0	20.9
Highly cong.	2	30.2		63.2		6.66		24.0	
Highly cong.	3	30.1	28.3	63.2	65.5	6.75	6.19	23.8	21.1
Highly cong.	3	30.0		63.2		6.76		23.8	

Finally, we analyze why the difference in social welfare becomes larger as the level of congestion increases. Although PMM users experience an increase in the cost of the trip due to the congestion toll, they are also benefited from a reduction in the travel time with respect to the base case. In the highly congested case, the contribution to the utility function of the travel time saving is somehow balanced out with the contribution associated with the increase in the cost (all else being equal). On the other hand, the waiting time of PT users increases in the three congestion situations because the frequency decreases from 6 trains/h to 5. However, this change is overcome by the decrease in the fare, which generates an increase in the utility of PT in the optimized case (all else being equal).

This experiment illustrates the flexibility of the choice-based optimization framework for the welfare maximization problem and its application to a realistic case study. Decomposition methods as the ones developed in Chapter 3 are relevant in this context because the populations in real-world transportation instances tend to be huge. Furthermore, a large number of draws is desired for the stabilization of the social welfare, as shown by the performed experiments.

4.5 Evaluation of a dedicated bus lanes policy

To the best of our knowledge, Basso and Silva (2014) appears to be the closest reference to what we propose. Although they present a non-linear optimization methodology, in this section we aim at investigating the strengths and limitations of our approach for their case study. We characterize a mixed-integer linear formulation for one of the policies that they assess based on the linear expressions introduced in Section 4.3. The authors rely on a nested logit model to capture the substitution between private and public transportation, inter-temporal and total demand elasticities. The model optimizes public transportation design variables in order to maximize a measure of social welfare, and it is used to evaluate the performance of alternative urban policies, such as congestion pricing or dedicated bus lanes and combinations thereof.

The formulation describing the tested policies is non-linear with continuously differentiable objective function and constraints. In order to solve it, they rely on a numerical algorithm from the software Wolfram Mathematica to find a local optimum of the problem. More precisely, a built-in function that searches for a local maximum with a given initial solution is used. As for the iterative method, the algorithm selects the best method for that computation among the available ones (e.g., conjugate gradient, quasi-Newton method).

The objective of this experiment is to provide a linearization scheme for the optimization problem associated with the dedicated bus lanes policy. The goal is not to linearize all the non-linear expressions in the optimization problem, as some simplifications are made for the sake of illustration, but to exemplify that the linear representation of the DCM, in addition to other linearization techniques, can be used to test urban policies in real-life contexts, and that similar conclusions can be made. Sections 4.5.1 and 4.5.2 describe the nested logit model and the non-linear optimization problem as defined in Basso and Silva (2014). Section 4.5.3 introduces the simplifications that we have made and the linear formulation that we propose, and Section 4.5.4 presents the numerical results.

4.5.1 Nested logit model

Basso and Silva (2014) model a representative kilometer of a urban road network where a bus service is offered, and the time horizon consists of one day of operation. The demand is composed of $N = 5$ groups of travelers (indexed by n), each of them representing an income group, and θ_n denotes the number of travelers per kilometer that belong to income group n. More specifically, $\theta_1 = 1960$, $\theta_2 = 3920$, $\theta_3 = 4480$, $\theta_4 = 2380$ and $\theta_5 = 1260$. The considered DCM is a nested logit model, whose nesting structure is depicted in Figure 4.6. The nests are defined according to the period the traveler aims at traveling (including a no-travel option to capture travelers leaving the system), and within each period both bus and car are considered by the individuals.

To refer to the nesting structure introduced by Basso and Silva (2014), we denote the nest by $q \in \mathcal{Q} = \{\text{peak}, \text{off-peak}, \text{no-travel}\}$ and the transportation mode by $i \in \mathcal{C} = \{\text{PMM}, \text{PT}\}$ (PMM represent the car and PT the bus). Note that this is equivalent to define $\mathcal{C}$ as the set of the alternatives defined by a combination of a period and a transportation mode and the no-travel option, and denote each alternative simply by i.

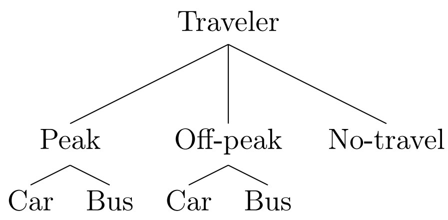

Figure 4.6: Decision tree associated with each individual (nesting structure)

The systematic term of the utility associated with period $q \in \mathcal{Q}' = \{\text{peak}, \text{off-peak}\}$, transportation mode $i \in \{\text{PMM}, \text{PT}\}$ and income group n is defined as

$$V_{qin} = \text{ASC}_{qin} + \beta_n^c c_{qi} + \beta_{qin}^t gt_{qi}, \qquad \forall q \in \mathcal{Q}', i \in \{\text{PMM}, \text{PT}\}, n \tag{4.39}$$

where ASC_{qin} is the alternative-specific constant, c_{qi} is the monetary cost of the trip, gt_{qi} is the generalized travel time, and β_n^c and β_{qin}^t are the cost and time parameters, respectively. Notice that both the cost and the travel time are assumed to be the same for all income groups. The systematic term of the no-travel alternative is set to a constant, i.e., $V_{\text{no-travel},n} = \text{ASC}_{\text{no-travel},n}, \forall n$. The utility functions are then defined as

$$U_{qin} = V_{qin} + \varepsilon_{qn} + \varepsilon_{qin}, \qquad \forall q \in \mathcal{Q}, i \in \{\text{PMM}, \text{PT}\}, n \tag{4.40}$$

where the nest-specific error terms ε_{qn} are independent and identically distributed (iid) $\text{EV}(0, \mu)$, and are the same for all alternatives within the nest, and $\varepsilon_{qin} \overset{iid}{\sim} \text{EV}(0, \mu_q)$ are the alternative-specific error terms. As determined by Basso and Silva (2014), $\mu = 0.25$ and the scale parameters at the lower level are normalized to 1, i.e., $\mu_q = 1, \forall q \in \mathcal{Q}$.

The estimates of the parameters are directly taken from Basso and Silva (2014), and are included in Table 4.6. Notice that all parameters but the cost are alternative-specific, i.e., the sensitivity towards cost is assumed to be the same regardless of the period and the mode.

The monetary cost of a bus trip is directly obtained from the bus fare, i.e.,

$$c_{q,\text{PT}} = p_{q,\text{PT}} \cdot \ell, \qquad \forall q \in \mathcal{Q}', \tag{4.41}$$

where $p_{q,\text{PT}}$ is the bus fare for period q [\$/km] and ℓ the average trip length [km]. In the case of car, its cost $c_{q,\text{PMM}}$ [\$/km] is equal to the congestion toll plus an operational cost (c_{oc}) related to expenses on fuel, tires, etc.:

$$c_{q,\text{PMM}} = (p_{q,\text{PMM}} + c_{\text{oc}})\frac{\ell}{\varrho}, \qquad \forall q \in \mathcal{Q}', \tag{4.42}$$

where ϱ is the average number of passengers in a car.

Table 4.6: Estimates of the parameters of the nested logit model defined by (4.39)–(4.40) (Basso and Silva, 2014)

Parameter	$n = 1$	$n = 2$	$n = 3$	$n = 4$	$n = 5$
$\text{ASC}_{\text{PMM,peak}}$	0	0	0	0	0
$\text{ASC}_{\text{PMM,off-peak}}$	−0.875	−1.01	−1.03	−1.11	−1.10
$\text{ASC}_{\text{PT,peak}}$	−1.10	−0.295	−0.870	−1.30	−2.18
$\text{ASC}_{\text{PT,off-peak}}$	−1.96	−1.68	−2.27	−2.56	−3.56
$\text{ASC}_{\text{no-travel}}$	−10.1	−8.73	−9.01	−8.79	−8.10
β^c	−1.69	−1.06	−0.929	−0.718	−0.422
$\beta^t_{\text{PMM,peak}}$	−0.655	−0.657	−0.676	−0.678	−0.656
$\beta^t_{\text{PMM,off-peak}}$	−0.478	−0.308	−0.314	−0.319	−0.190
$\beta^t_{\text{PT,peak}}$	−1.13	−1.13	−1.16	−1.17	−1.13
$\beta^t_{\text{PT,off-peak}}$	−1.19	−0.769	−0.785	−0.796	−0.475

The generalized travel time by bus [h] in period q (per kilometer) when dedicated bus lanes are in place is given by

$$gt_{q,\text{PT}} = t_{q,\text{PT}}\ell + \phi_1 t_q^{\text{w}} + \phi_2 t_{\text{acc}}, \qquad \forall q \in \mathcal{Q}', \tag{4.43}$$

where $t_{q,\text{PT}}$ is the in-vehicle travel time per kilometer in period q (described below), ℓ is the average trip length, t_q^{w} is the waiting time at the bus stop, t_{acc} is the access time, defined as the time to walk to and from the bus stop, and ϕ_1 and ϕ_2 are the weights capturing the dislike of travelers towards waiting and walking times, respectively. The waiting time t_q^{w} is a fraction ν of the interval between buses, i.e.,

$$t_w^q = \frac{\nu}{f_q}, \qquad \forall q \in \mathcal{Q}', \tag{4.44}$$

where f_q is the bus frequency [bus/h] in period q. The access time t_{acc} is calculated from the average walking distance to access the bus stop ($1/4s$), where s represents the number of equidistant stops per kilometer. Hence, as the average walking distance to access the bus stop is the same as the average distance from the bus stop to the destination, the access time is obtained as

$$t_{\text{acc}} = \frac{1}{2sV_{\text{w}}}, \tag{4.45}$$

where V_{w} is the walking speed.

The in-vehicle travel time by bus in period q (per kilometer) in the case of dedicated bus lanes is given by

$$t_{q,\text{PT}} = t_f\left(1 + a\left(\frac{f_q \upsilon(k_{\text{PT}})}{mC}\right)^b\right) + s_{\text{PT}}\left(\frac{D_{q,\text{PT}}}{H_q f_q s_{\text{PT}}} t_{\text{sb}} + t_{\text{d}}\right), \qquad \forall q \in \mathcal{Q}'. \tag{4.46}$$

The first term on the right-hand side corresponds to the time that a bus spends while in motion, where t_f is the free-flow travel time, a and b are the parameters of the BPR function (see (4.14)), k_{PT} represents the bus capacity [travelers/bus] and $v(k_{\mathrm{PT}})$ is a linear function on k_{PT} providing an equivalence factor between buses and cars which increases with k_{PT}, C is the capacity of the road [cars/h] and m denotes the fraction of C dedicated to bus lanes. The second-term on the right-hand side is the time spent at bus stops, and it is given by the number of stops each bus makes in each kilometer (s_{PT}) multiplied by the time spent at each bus stop. The number of passengers boarding at each bus stop is $D_{q,\mathrm{PT}}/(H_q f_q p)$, where $D_{q,\mathrm{PT}}$ is the bus demand at period q, H_q is the length (in hours) of the time period, f_q is the bus frequency and t_{sb} is the time each passenger takes to board a bus. Finally, t_{d} is a non-linear function representing bus congestion at the bus stop, i.e., buses queuing to get in and out of the bus stop (this function is not specified in Basso and Silva, 2014 and has been extracted from the code used by the authors for their experiments).

The generalized travel time for car in period q [h] is given only by the in-vehicle travel time

$$gt_{q,\mathrm{PMM}} = t_{q,\mathrm{PMM}} \cdot \ell, \tag{4.47}$$

where $t_{q,\mathrm{PMM}}$ is the travel time per kilometer. It is calculated as

$$t_{q,\mathrm{PMM}} = t_f \left(1 + \alpha \left(\frac{\ell D_{q,\mathrm{PMM}}/(H_q a)}{(1-m)C} \right)^{\beta} \right), \qquad \forall q \in \mathcal{Q}', \tag{4.48}$$

where $\ell \cdot D_{q,\mathrm{PMM}}/(H^q \varrho)$ is the car flow [cars/h], as $D_{q,\mathrm{PMM}}$ is the car demand at period q per kilometer, H_q the period duration in hours, ℓ is the average trip length and ϱ is the (constant) car occupancy. Notice that the road capacity (number of lanes) available to cars is obtained as $(1-m)C$ (m is the fraction of lanes dedicated to buses).

4.5.2 Optimization problem

The planner can optimize the design variables of the bus system: frequency f_q [bus/h] in each period $q \in \mathcal{Q}'$, bus capacity k_{PT} [travelers/bus] and number of equidistant bus stops per kilometer s_{PT}. Notice that these variables are discrete in nature, but are defined as continuous variables in Basso and Silva (2014). The planner can also set the prices: the bus fare for each period $p_{q,\mathrm{PT}}$ [$/km] and the congestion toll for cars for each period $p_{q,\mathrm{PMM}}$ [$/km]. Another decision that is included in the problem is the fraction m of the capacity that is exclusively dedicated to bus lanes.

The objective function to be maximized is an unweighted social welfare defined for one kilometer of a day of operation. It includes the consumer surplus and the difference between the revenues obtained from congestion tolls and bus fares and the implementation costs of the policy in place.

As described in Section 4.3.1, the consumer surplus in the nested logit model is calculated using formula (4.5) with CPGF (4.6), and can be expressed in monetary units as follows:

$$\mathrm{CS} = \sum_{n=1}^{N} \left(\frac{1}{\mu} \frac{\theta_n}{-\beta_n^{\mathrm{c}}} \ln \sum_{q \in \mathcal{Q}} \mathrm{e}^{\mu A_n(q)} \right), \tag{4.49}$$

where μ is the upper-level scale parameter of the nested logit model, θ_n is the size of income group n, β_n^c is the cost sensitivity parameter associated with income group n, and $A_n(q)$ are the expected utilities of the nests, defined as $A_{\text{no-travel},n} = \mathrm{ASC}_{\text{no-travel},n}, \forall n$ for the no-travel nest, and with the following expression for the other nests:

$$A_n(q) = \ln \sum_{j \in \{\mathrm{PMM}, \mathrm{PT}\}} \mathrm{e}^{V_{qjn}}, \qquad \forall q \in \mathcal{Q}'. \tag{4.50}$$

The revenue obtained from bus fares and congestion tolls is given by:

$$\mathrm{G} = \sum_{q \in \mathcal{Q}'} D_{q,\mathrm{PT}} \cdot p_{q,\mathrm{PT}} \cdot \ell + \sum_{q \in \mathcal{Q}'} D_{q,\mathrm{PMM}} \cdot p_{q,\mathrm{PMM}} \frac{\ell}{\varrho}(1-\kappa), \tag{4.51}$$

where κ is a fraction of the car revenue that represents the implementation costs of congestion pricing.

The operating costs of the bus system per day per kilometer are calculated as a function of the bus fleet (B_{PT}), the total number of vehicle-kilometers of each period (Λ_q), and the bus size (k_{PT}) as follows:

$$\mathrm{OC} = \mathrm{OC}_B(k_{\mathrm{PT}}) B_{\mathrm{PT}} + \sum_{q \in \mathcal{Q}'} \mathrm{OC}_\Lambda(k_{\mathrm{PT}}) \Lambda_q. \tag{4.52}$$

The first term of the right-hand side mainly corresponds to labor and vehicle-capital expenses, and the second term captures operational expenses. The required fleet of vehicles is defined as $B_{\mathrm{PT}} = \max_{q \in \mathcal{Q}'} \{f_q \cdot t_{q,\mathrm{PT}}\}$, and it is assumed that the peak period is the one that characterizes the amount of buses required for operation, i.e., $B_{\mathrm{PT}} = f_{\mathrm{peak}} \cdot t_{\mathrm{peak},\mathrm{PT}}$. The daily number of vehicles per kilometer is obtained as $\Lambda_q = H_q f_q$ for $q \in \mathcal{Q}'$. The functions $\mathrm{OC}_B(k_{\mathrm{PT}})$ and $\mathrm{OC}_\Lambda(k_{\mathrm{PT}})$ are linear functions that respectively transform B_{PT} and $\Lambda_q, \forall q \in \mathcal{Q}'$, into monetary units.

The final expression for the social welfare is

$$\mathrm{SW} = \mathrm{CS} + \mathrm{mcpf}\left(\mathrm{G} - \mathrm{OC} - \mathrm{OC}_{\mathrm{dl}}\right), \tag{4.53}$$

where $\mathrm{OC_{dl}}$ is the (constant) cost of implementing and operating dedicated bus lanes (with respect to the mixed traffic case), and mcpf > 1 refers to the marginal cost of public funds, which are considered in Basso and Silva (2014) to capture the fact that such funds are costly, i.e., this term works as a proxy for cost inefficiencies induced by subsidies.

The frequency f_q is constrained to be positive and less than the capacity of bus stops for period q. The capacity of a bus stop is a function of boarding and alighting times that depend on many of the optimization variables. The bus fare $p_{q,\mathrm{PT}}$ and the congestion toll $p_{q,\mathrm{PMM}}$ are restricted to be positive in both periods. In this application, as the road capacity is equal to three lanes, the fraction m of the capacity dedicated to bus lanes can only be one or two thirds (there has to be at least one lane for cars). The bus capacity k_{PT} is equal to the passenger load because the crowd dislike by travelers is not incorporated in the nested logit model and a larger capacity will worsen the objective function. Hence, the bus size is calculated as

$$k_{\mathrm{PT}} = \max_{q \in \mathcal{Q}'} \frac{D_{q,\mathrm{PT}} \cdot \ell}{f_q H_q}. \tag{4.54}$$

We denote by ρ a fixed percentage of subsidization of the bus system. The following constraint imposes the cost of the bus system that is not subsidized to be self-financed:

$$\sum_{q \in \mathcal{Q}'} D_{q,\mathrm{PT}} p_{q,\mathrm{PT}} \ell = \mathrm{OC}(100 - \rho)/100. \tag{4.55}$$

Notice that $\rho = 0$ means that no subsidization is assumed.

4.5.3 Linear formulation

In order to simplify the approach, we assume that the bus capacity k_{PT} and the number of equidistant stops per kilometer s_{PT} are fixed to the optimal values obtained in Basso and Silva (2014). We provide the values in Section 4.5.4 as they are specific to the two situations being tested. Furthermore, as the road capacity of the considered application has only three lanes, the fraction m of the lanes dedicated exclusively to buses can only take two values (1/3 or 2/3), so we also fix its value to the reported optimal value (1/3). Notice that it is possible to include these decisions as binary variables in the formulation. Hence, as in Section 4.4, the design variables are the congestion toll, the PT fare and the frequency, i.e., $p_{q,\mathrm{PMM}}$, $p_{q,\mathrm{PT}}$ and f_q, $\forall q \in \mathcal{Q}'$. As the capacity of the modes is not explicitly considered in Basso and Silva (2014), we formulate the problem without capacity constraints.

For the linear formulation of the systematic parts of the utility functions (4.39), we need to determine linear expressions on the travel time and cost as functions of the endogenous variables and the demand. The in-vehicle travel time by bus is linearly characterized for each iteration k of the iterative approach described in Algorithm 4.1 with the linear representation of the frequency (see (4.1)–(4.2)) and the demand $D^k_{q,\mathrm{PT}}$ as

$$\begin{aligned} t^k_{q,\mathrm{PT}} =& t_f\left(1 + a\left(\frac{\upsilon(k_{\mathrm{PT}})}{mC}\right)^b \left(\sum_{\ell=1}^{L} (f_{\ell q})^b \tau_{\ell q}\right)\right) \\ &+ \frac{D^k_{q,\mathrm{PT}}}{H_q s_{\mathrm{PT}}}\left(\sum_{\ell=1}^{L}\frac{1}{f_{\ell q}}\tau_{\ell q}\right) t_{\mathrm{sb}} + s_{\mathrm{PT}} \cdot t^k_d, \qquad \forall q \in \mathcal{Q}', \end{aligned} \tag{4.56}$$

where the demand $D^k_{q,\mathrm{PT}}$ is calculated with the choice variables obtained in the previous iteration $\bar{w}^{k-1}_{qinr}$, i.e.,

$$D^k_{qi} = \frac{1}{R}\sum_{n=1}^{N}\sum_{r=1}^{R}\theta_n \bar{w}^{k-1}_{qinr}, \qquad \forall q \in \mathcal{Q}', i \in \{\mathrm{PMM}, \mathrm{PT}\}. \tag{4.57}$$

Notice that the only decision variable that appears in (4.56) is $\tau_{\ell q}$, which are the binary variables that define the frequency associated with period q (f_q). The remaining elements are either constant values or are fixed to the values obtained at the previous iteration of the iterative approach. This is the case of the non-linear function t_d, that depends on the frequency f_q and the demand $D_{q,\mathrm{PT}}$. Thus, instead of fixing this function to some given value, we update t_d at each iteration by updating the values of the corresponding variables.

As the waiting time for the bus alternative also depends on the frequency, the linear representation of (4.44) is given by

$$t^q_w = \sum_{\ell=1}^{L}\frac{\nu}{f_{\ell q}}\tau_{\ell q}, \qquad \forall q \in \mathcal{Q}'. \tag{4.58}$$

Similar to $t^k_{q,\mathrm{PT}}$, the travel time by car (4.48) is linearized by

$$t^k_{q,\mathrm{PMM}} = t_f\left(1 + \alpha\left(\frac{\ell D^k_{q,\mathrm{PMM}}/(H_q a)}{(1-m)C}\right)^\beta\right), \qquad \forall q \in \mathcal{Q}'. \tag{4.59}$$

As the demand $D^k_{q,\mathrm{PMM}}$ is provided by the iterative algorithm (given by (4.57)), this expression does not contain any decision variable of the linear formulation.

With respect to the simulation draws, we note that in the nested logit model it is not possible to draw directly from the random term because of the correlation structure between the alternatives. As is done in Bortolomiol et al. (2019), we exploit the logit-like formulation of MEV models as an approximation. We define $V'_{qin}, \forall q \in \mathcal{Q}', i \in$

$\{\text{PMM}, \text{PT}\}, n$ as

$$V'_{qin} = V_{qin} + \log\left(\mu \exp(V_{qin}(\mu_q - 1))\left(\sum_{j \in M_q} \exp(V_{qjn}\mu_m)^{\frac{\mu}{\mu_q}-1}\right)\right), \tag{4.60}$$

where M_q represents the set of alternatives included in nest q, μ is the scale parameter associated with the upper-level nest, and μ_q is the scale parameter associated with the lower-level nest (as described in Section 4.5.1). This characterization of the systematic term of the utility allows us to draw from an extreme value distribution as in a logit model.

Expression (4.60) is non-linear due to the presence of the endogenous variables $p_{q,\text{PT}}$, $p_{q,\text{PMM}}$ and f_q (which are present in V_{qin} via the travel cost and the travel time). Thanks to Algorithm 4.1, we can overcome this issue by fixing the values of the endogenous variables in the logsum term (first to some initial values and then to the values reported by the optimal solution at the previous iteration) and only optimize for those in V_{qin}.

For the measure of social welfare, the consumer surplus (4.49) expressed in monetary units is calculated as follows:

$$\text{CS} = \frac{1}{R}\sum_{n=1}^{N} \frac{\theta_n}{-\beta_n^c} \sum_{r=1}^{R} U_{nr}, \tag{4.61}$$

The revenue obtained from bus fares and congestion tolls is calculated with the variables η_{qinr}, i.e.,

$$\text{G} = \sum_q \sum_{n=1}^{N} \sum_{r=1}^{R} \eta_{q,\text{PT},nr} \cdot \ell + \sum_q \sum_{n=1}^{N} \sum_{r=1}^{R} \eta_{q,\text{PMM},nr} \cdot \frac{\ell}{\varrho}(1-\kappa). \tag{4.62}$$

Finally, the operating costs (4.52) are obtained by calculating the fleet of vehicles B_{PT} as the product $f_{\text{peak}} \cdot t^k_{\text{peak,PT}}$, which leads to a linear expression thanks to characterization of f_q via binary variables, and the daily number of vehicles per kilometer Λ_q, which is already a linear expression on f_q.

With respect to the constraints, as the bus size k_{PT} is assumed to be given, we express (4.54) as an inequality constraint on f_q, i.e.,

$$f_q \geq \frac{D_{q,\text{PT}} \cdot \ell}{k_{\text{PT}} H_q}, \qquad \forall q \in \mathcal{Q}', \tag{4.63}$$

where D_{qi} is obtained as in (4.57) with the current choice variables, i.e.,

$$D_{qi} = \frac{1}{R}\sum_{n=1}^{N}\sum_{r=1}^{R} \theta_n w_{qinr}, \qquad \forall q \in \mathcal{Q}', i \in \{\text{PMM}, \text{PT}\}. \tag{4.64}$$

Constraint (4.55) is linearly expressed by considering the expression on the revenue obtained from bus fares (first term in (4.62)) and the linear representation of the cost OC previously described. The linear formulation of the optimization problem is included in Model 4.7 (we denote by a_{qi} and b_{qi} the lower and upper bound on p_{qi}, respectively).

$$
\begin{aligned}
SW(D^k) = \max \quad & \mathrm{CS} + \mathrm{mcpf}\,(\mathrm{G} - \mathrm{OC} - \mathrm{OC_{dl}}) && && (4.65)\\
\text{s.t.} \quad & U_{qinr} = V'_{qin} + \xi_{qinr} && \forall i \in \mathcal{C}, q \in \mathcal{Q}, n, r && (4.66)\\
& V'_{qin} = V_{qin} + \log\left(\mu \exp(V_{qin}(\mu_q - 1))\left(\sum_{j\in M_q} \exp(V_{qjn}\mu_m)^{\frac{\mu}{\mu_q}-1}\right)\right) && \forall i \in \mathcal{C}, q \in \mathcal{Q}, n && (4.67)\\
& V_{qin} = \mathrm{ASC}_{qin} + \beta^c_n c_{qi} + \beta^t_{qin} gt^k_{qi} && \forall i \in \mathcal{C}, q \in \mathcal{Q}, n && (4.68)\\
& c_{q,\mathrm{PMM}} = (p_{q,\mathrm{PMM}} + c_{\mathrm{oc}})\frac{\ell}{\varrho} && \forall q \in \mathcal{Q}' && (4.69)\\
& c_{q,\mathrm{PT}} = p_{q,\mathrm{PT}} \cdot \ell && \forall q \in \mathcal{Q}' && (4.70)\\
& gt^k_{q,\mathrm{PMM}} = t^k_{q,\mathrm{PMM}} \cdot \ell && \forall q \in \mathcal{Q}' && (4.71)\\
& gt^k_{q,\mathrm{PT}} = t^k_{q,\mathrm{PT}} \cdot \ell + \phi_1 t^{\mathrm{w}}_q + \phi_2 t_{\mathrm{acc}} && \forall q \in \mathcal{Q}' && (4.72)\\
& t^k_{q,\mathrm{PT}} = t_f\left(1 + a\left(\frac{\upsilon(k_{\mathrm{PT}})}{mC}\right)^b\left(\sum_{\ell=1}^{L_f}(f_{\ell q})^b \tau_{\ell q}\right)\right) + \frac{D^k_{q,\mathrm{PT}}}{H_q s}\left(\sum_{\ell=1}^{L_f}\frac{1}{f_{\ell q}}\tau_{\ell q}\right) t_{\mathrm{sb}} + s_{\mathrm{PT}} \cdot t^k_d && \forall q \in \mathcal{Q}' && (4.73)\\
& t^q_w = \sum_{\ell=1}^{L_f} \frac{\nu}{f_{\ell q}}\tau_{\ell q} && \forall q \in \mathcal{Q} && (4.74)\\
& t^k_{q,\mathrm{PMM}} = t_f\left(1 + \alpha\left(\frac{\ell D^k_{q,\mathrm{PMM}}/(H_q a)}{(1-m)C}\right)^\beta\right) && \forall q \in \mathcal{Q}' && (4.75)\\
& f_q \geq \frac{\frac{1}{R}\sum_{n=1}^N \sum_{r=1}^R \theta_n w_{q,\mathrm{PT},nr} \cdot \ell}{k_{\mathrm{PT}} H_q} && \forall q \in \mathcal{Q}' && (4.76)\\
& U_{qinr} \leq U_{nr} && \forall i \in \mathcal{C}, q \in \mathcal{Q}, n, r && (4.77)\\
& U_{nr} \leq U_{qinr} + M_{qinr}(1 - w_{qinr}) && \forall i \in \mathcal{C}, q \in \mathcal{Q}, n, r && (4.78)\\
& \sum_{i\in C}\sum_{q\in\mathcal{Q}} w_{qinr} = 1 && \forall n, r && (4.79)\\
& a_{qi} w_{qinr} \leq \eta_{qinr} && \forall i \in \mathcal{C}, q \in \mathcal{Q}', n, r && (4.80)\\
& \eta_{qinr} \leq b_{qi} w_{qinr} && \forall i \in \mathcal{C}, q \in \mathcal{Q}', n, r && (4.81)\\
& p_{qi} - (1 - w_{qinr}) b_{qin} \leq \eta_{qinr} && \forall i \in \mathcal{C}, q \in \mathcal{Q}', n, r && (4.82)\\
& \eta_{qinr} \leq p_{qi} - (1 - w_{inr}) a_{qi} && \forall i \in \mathcal{C}, q \in \mathcal{Q}', n, r && (4.83)\\
& f_q = \sum_{\ell=1}^{L_f} f_\ell \tau_{\ell q} && \forall q \in \mathcal{Q}' && (4.84)\\
& \mathrm{CS} = \frac{1}{R}\sum_{n=1}^N \frac{\theta_n}{-\beta^c_n}\sum_{r=1}^N U_{nr} && && (4.85)\\
& \mathrm{G} = \sum_{q\in\mathcal{Q}'}\sum_{n=1}^N\sum_{r=1}^R \eta_{\mathrm{PT},qnr} \cdot \ell + \sum_{q\in\mathcal{Q}'}\sum_{n=1}^N\sum_{r=1}^R \eta_{\mathrm{PMM},qnr} \cdot \frac{\ell}{\varrho}(1-\kappa) && && (4.86)\\
& \mathrm{OC} = \mathrm{OC}_B(k_{\mathrm{PT}})B_{\mathrm{PT}} + \sum_{q\in\mathcal{Q}'} \mathrm{OC}_\Lambda(k_{\mathrm{PT}})\Lambda_q && && (4.87)\\
& \sum_{q\in\mathcal{Q}'} D_{q,\mathrm{PT}} p_{q,\mathrm{PT}} \ell = \mathrm{OC}(100 - \rho)/100 && && (4.88)\\
& B_{\mathrm{PT}} = f_{\mathrm{peak}} \cdot t^k_{\mathrm{peak,PT}} && && (4.89)\\
& w_{qinr}, \tau_\ell \in \{0,1\}, && \forall i \in \mathcal{C}, \forall q \in \mathcal{Q}, n, r, \ell && (4.90)
\end{aligned}
$$

Model 4.7: Linear formulation of the dedicated bus lanes policy (Basso and Silva, 2014)

4.5.4 Numerical results

In this section, we test the linear formulation introduced in Section 4.5.3 in the context of dedicated bus lanes with the simulated data from the city of Santiago (Chile) used in Basso and Silva (2014). As is done in Section 4.4, we define a base case and an optimized case and compare the corresponding difference in consumer surplus with the one obtained in Basso and Silva (2014).

The base case consists on the dedicated bus lanes policy without congestion pricing, i.e., $p_{q,\text{PMM}} = 0, \forall q \in \{\text{peak}, \text{off-peak}\}$, and the optimized case adds to the dedicated bus lanes policy the congestion pricing and the public transportation subsidization. We use the same notation as in Basso and Silva (2014) and denote the policies by DL and DL+CON+SUB, respectively. In both cases, the bus fare is assumed to be the same for both periods, i.e.,

$$p_{\text{peak},\text{PT}} = p_{\text{off-peak},\text{PT}}. \tag{4.91}$$

As mentioned in Section 4.5.3, the variables k_{PT} (bus capacity) and s_{PT} (number of equidistant stops per kilometer) are fixed to the optimal values obtained in Basso and Silva (2014). More precisely, $k_{\text{PT}} = 152.9$ and $s_{\text{PT}} = 3.04$ in DL and $k_{\text{PT}} = 159.2$ and $s_{\text{PT}} = 3.08$ in DL+CON+SUB.

We consider Algorithm 4.1 with the flows of PMM and PT for both periods and the endogenous variables $p_{q,\text{PT}}$, $p_{q,\text{PMM}}$ and f_q initialized to the initial values considered by the authors for their numerical experiments. We also consider the same subsidy for the DL+CON+SUB policy ($\rho = 55$) and the same price bounds for the congestion toll and the bus fare, i.e., $p_{q,\text{PMM}}, p_{q,\text{PT}} \in [0, 0.02], \forall q \in \{\text{peak}, \text{off-peak}\}$. We set $L = 6$ frequency levels for both periods, with $f_{\text{peak}} \in \{41, 42, 43, 44, 45, 46\}$ and $f_{\text{off-peak}} \in \{13, 14, 15, 16, 17, 18\}$.

Figure 4.8 compares the difference in consumer surplus between DL and DL+CON+SUB for the 3 considered replications and different values of R. As the number of individuals is small ($N = 5$), it is possible to test the problem with a larger number of draws than in Section 4.4 for a similar computational time. Furthermore, given the large sizes of the groups, a higher variability across simulation draws is expected, and therefore larger values of R might be required. The reference line indicates the difference in consumer surplus obtained in Basso and Silva (2014).

We can observe that the linear approximation of the consumer surplus gets closer to the reference value as the number of draws increases, as expected. Even though we obtain a rather low difference for one of the replications for $R = 250$, the results are in line with the reference value, especially for $R = 500$ and $R = 1000$.

Table 4.7 shows the average computational time of Algorithm 4.1 for the problem in Model 4.7 for three replications. As the population is composed of only 5 individuals (each of them representing a group with homogenous behavior), the computational times are expected to be low. The increase in computational time from DL to DL+CON+SUB is remarkable, as the average computational time rises from 1.25 min to 17 min for $R = 500$ and from 3 min to 1.8 h for $R = 1000$.

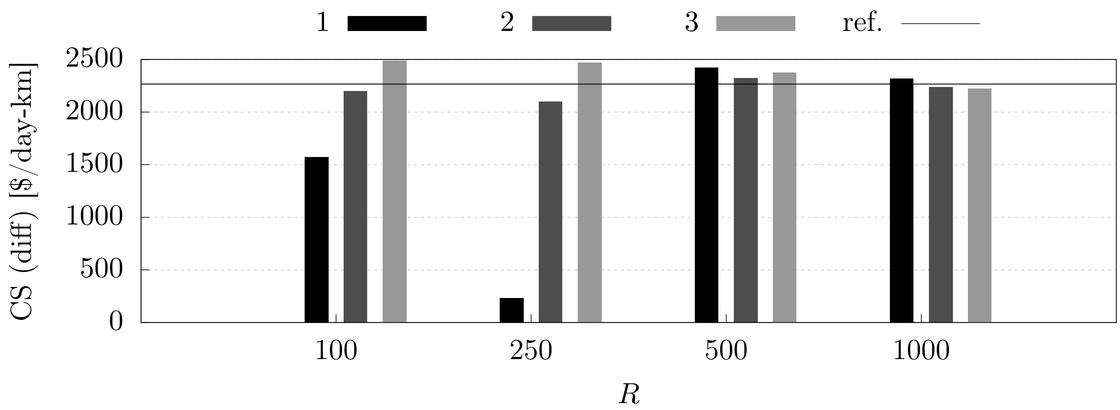

Figure 4.8: Difference in consumer surplus [$/day-km] between DL and DL+CON+SUB for the 3 considered replications (independent generation of draws, labeled as 1-3) for $R \in \{100, 250, 500, 1000\}$ (ref. represents the value reported by Basso and Silva (2014))

Table 4.7: Average computational time (min) of Algorithm 4.1 for 3 replications (independent generation of draws) for the problem in Model 4.7 and the two evaluated policies

R	DL	DL+CON+SUB
100	0.103	0.505
250	0.922	3.47
500	1.25	16.9
1000	2.76	106.0

Table 4.8 includes the reference flows [veh/h] and the range of flows defined by the different solutions obtained with the three replications for the linear formulation and $R = 1000$. Similarly, Table 4.9 presents the reference values of the design variables (toll, fare and frequency) and the range of values defined by the different solutions obtained with the 3 considered replications for $R = 1000$. Notice that the toll is not defined in the DL policy for any of the periods.

We observe that the flows are approximated by the linear formulation, with a slight tendency in the off-peak period for narrower ranges and closer values to the reference values. Even though the DL+CON+SUB policy allows for a congestion toll, it is set to 0 by both models. The bus fare is also reproduced by the linear formulation, and the frequency levels reported by the different solutions are in line with the continuous reference ones.

Finally, Figure 4.9 illustrates a distributional analysis with respect to the 5 income groups, which are labeled from the lowest to the highest. We observe that the lower the income, the higher the benefit obtained by the traveler from the implementation of DL+CON+SUB with respect to DL, which is expected.

We notice that the linear formulation here proposed allows to incorporate decisions that are discrete in nature as integer variables, i.e., we do not need to assume that such variables are real-valued in order to derive the formulation. Furthermore, the methodology we present allows to accommodate any DCM, whereas Basso and Silva (2014) rely on the closed-form expression of the consumer surplus of the nested logit model. Although not explored in this experiment, additional assumptions such as capacity requirements on a transportation mode could be incorporated in the linear formulation.

Table 4.8: Comparison of the range of flows [veh/h] obtained with 3 replications (independent generation of draws) with the results obtained in Basso and Silva (2014) for $R = 1000$

Mode	Period	DL		DL+CON+SUB	
		Reference	Linear	Reference	Linear
PMM	peak	380.7	[376.3,394.3]	352.8	[346.9,357.7]
PMM	off-peak	196.5	[198.5,201.6]	174.3	[172.0,174.4]
PT	peak	643.6	[614.2,641.2]	679.3	[652.4,681.3]
PT	off-peak	241.1	[236.9,249.9]	262.7	[260.2,274.7]

Table 4.9: Comparison of the range of values of the design variables obtained with 3 replications (independent generation of draws) with the results obtained in Basso and Silva (2014) for $R = 1000$

Mode	Period	Variable	DL		DL+CON+SUB	
			Ref.	Lin.	Ref.	Lin.
PMM	peak	toll [$/km]	NA	NA	0	0
PMM	off-peak	toll [$/km]	NA	NA	0	0
PT	peak	fare [$/km]	0.047	[0.0466,0.0473]	0.021	[0.0206,0.021]
PT	peak	freq. [bus/h]	42.1	{41,42}	42.7	{41,42,43}
PT	off-peak	fare [$/km]	0.047	[0.0466,0.0473]	0.021	[0.0206,0.021]
PT	off-peak	freq. [bus/h]	15.8	{16,17}	16.5	{17,18}

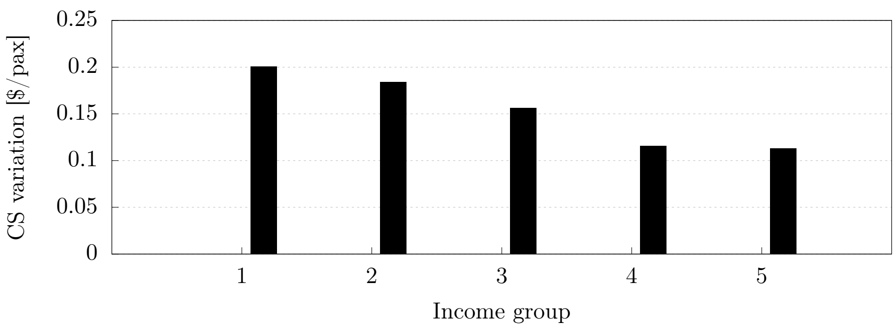

Figure 4.9: Average variation in consumer surplus [$/pax] between DL and DL+CON+SUB for the 3 considered replications (independent generation of draws) for $R = 1000$

4.6 Concluding remarks

In this chapter, we show that the choice-based optimization framework can accommodate other types of optimization problems. We address the problem of pricing and design of a transportation system such that a measure of social welfare is maximized. Thanks to the linear representation of DCM, an approximation of the consumer surplus, which defines the social welfare, is indeed readily available from our formulation. We also propose linear formulations of other relevant features in this context: a budgetary constraint that enables to incorporate a revenue recycling mechanism and a formulation of the travel time that depends on the demand in order to capture congestion effects.

As the travel time is assumed to depend on the demand, and travel time is one of the variables explaining the choice of individuals (and therefore determines the demand), there is a fixed-point problem. In order to test the methodology, we define an iterative algorithm for the resulting optimization problem, and we develop two different case studies. The former is a semi-synthetic case study that considers an Integrated Choice and Latent Variable (ICLV) model to explain the behavior of users. The objective is to evaluate the effect of the introduction of a congestion toll on the modal split and the other design variables of the system with respect to its current state. The latter consists of a linearization scheme for one of the policies being tested in a real-life case study from the literature that relies on a highly non-linear formulation. The objective is to provide general guidelines on the usage of the proposed methodology in such applications.

The described modeling framework allows to decide on the features of a transportation system while accounting for the impact of such decisions on the social welfare and modal split, which enables the evaluation of different policies with respect to the criteria established by the transportation authority. In Section 4.4, the introduction of a toll for PMM leads to an increase of the consumer surplus and a modal shift from PMM to PT. In Section 4.5, the same outcome is obtained by introducing a PT subsidization to the dedicated bus lanes policy.

The conducted experiments show that a larger number of draws is required for the objective function to stabilize (in comparison with the profit maximization problem, for instance), which is expected because of the definition of the consumer surplus via the expected maximum utilities. Nevertheless, in the absence of capacity constraints, and thanks to the implementation of additional strategies such as the grouping of individuals, it is possible to enlarge the number of draws without inducing an excessive computational burden. Furthermore, the Lagrangian decomposition scheme developed in Section 3.5 could be adapted to handle the welfare-maximization optimization problem for larger instances, and/or when additional modeling features are included, such as capacity constraints.

The methodology explained in this chapter provides a general framework for the evaluation of alternative transportation policies with respect to a measure of social welfare and distributional analyses. It allows to assess the nature of a policy (regressive or progressive), and to identify the segments in the population that are most adversely impacted. Moreover, the model offers a great flexibility when it comes to incorporate other features of a transportation system, such as revenue recycling mechanisms, and to accommodate advanced DCM that were typically not considered because of the complexity of the associated welfare measures.

5 Conclusion

5.1 Main findings and implications

This thesis proposes a general modeling framework for the integration of discrete choice models (DCM) in mixed-integer linear problems (MILP) in order to provide a disaggregate demand representation that captures the interplay between the behavior of individuals and the supply-related decisions to be optimized. The introduced methodology is able to keep the sophistication of DCM while ensuring that the resulting formulation remains operational. Furthermore, the development of algorithms that exploit the decomposable structure of the model enables to deal with its complexity in order to enhance its tractability. The applications explored in this thesis illustrate not only the relevance of accounting for the interactions between the demand and the supply actors, but also the flexibility of the approach when it comes to accommodate distinct modeling features in a variety of different contexts.

Chapter 2 introduces a choice-based optimization framework for the inclusion of advanced DCM into MILP. We have proposed a formulation that relies on simulation to express the behavioral preference structure of individuals directly in terms of the utility functions (instead of the associated probability expressions), which leads to a set of mixed-integer linear constraints that can be embedded in any MILP formulation. The interaction between the demand and the supply-related decisions is modeled with the so-called endogenous variables. They represent the decisions that have an impact on the behavior of individuals, and must linearly appear on the utility functions for their insertion in the MILP model.

By means of the revenue maximization problem, the problem of an operator that wants to set a pricing strategy such that its revenue is maximized, we show how the modeling framework can be employed. We are able to include capacity constraints on the services being offered and provide alternative formulations for the pricing strategy to be proposed to the individuals. The case study considered as proof-of-concept shows how an advanced DCM from the literature can be integrated in the revenue maximization problem as such. It also sheds light on the versatility of the framework by testing different aspects of the formulation, such as price differentiation by market segmentation, the grouping of individuals with homogeneous behavior and the maximization of the profit subject to a capacity allocation strategy. These experiments also exhibit the computational complexity prompted by the disaggregate demand representation and the simulation-based linearization of the modeling framework.

Motivated by the fact that these two aspects represent two differentiated dimensions of the model (the individuals and the simulation draws), in Chapter 3 we rely on decomposition techniques to exploit the decomposable structure of the mathematical formulation. Again, the revenue maximization problem is considered to characterize the strategies being explored. Various complicating constraints (i.e., utility functions) that link blocks of the formulation can be identified. Nevertheless, Lagrangian relaxation on these constraints turned out to be unsuitable due to the strong interrelations between the variables and constraints in the model. In order to overcome this issue, we develop a heuristic approach based on Lagrangian decomposition that allows to preserve all the original constraints within the subproblems. We induce decomposition of the formulation by gathering simulation draws into subsets and generating duplicates of the price variables for each subset. The grouping of simulation draws is proposed to decrease the number of introduced price duplicates (instead of one price duplicate per draw), which we show has an impact on the convergence of the method.

This approach is enclosed in the subgradient method, so that at every iteration the Lagrangian subproblems are solved and feasible solutions of the original optimization problem are generated. The performed tests show that near-optimal solutions are obtained in a much reduced computational time (by running only 10% of the computational time used by the exact method). Moreover, we observe that as long as the subproblems remain computationally manageable, a large number of draws per subset is recommended, as it leads to smaller duality gaps for a given running time.

Chapter 4 presents an alternative application of the choice-based optimization framework in the context of welfare maximization in transportation, where the role of the operator is taken by a public transportation authority. The key quantitative element of welfare analysis in the context of DCM, the expected maximum utility, is readily available in the model, as it is required to determine the choices of individuals. Thus, it can be easily included in the objective function to define a measure of social welfare. This is a significant advantage because it enables not to deal with the complex non-linear formulations of this quantity as provided by discrete choice theory.

We characterize a model for the pricing and design of a transportation system in the presence of a revenue recycling mechanism such that the welfare is maximized. The decision variables of the public transportation authority (road price and public transportation fare and frequency) are endogenous variables, as they have a direct impact on the modal split. We test the formulation on two different case studies. The former evaluates the effect of the introduction of a congestion toll in an urban setting on the modal split and social welfare, and the latter provides a linearization scheme for a dedicated bus lanes policy that is tested in a real-life case study from the literature that relies on a highly non-linear formulation. In comparison with the revenue maximization problem, the conducted experiments show that a larger number of draws is required for the objective function to stabilize, which is expected due to the definition of the objective function via the expected maximum utility. This can be addressed with some of the strategies previously introduced (e.g., grouping of individuals) and by adapting the Lagrangian decomposition scheme to this problem.

In summary, from a theoretical point of view, we have introduced an operational modeling framework that enables the integration of DCM in MILP formulations to account for the interactions between the demand and the supply. We have also proposed an algorithmic approach based on Lagrangian decomposition that aims at decreasing the computational burden associated with the modeling framework while being able to generate feasible solutions with a low optimality gap. From a practical point of view, we have illustrated the extent of the framework by extensively analyzing two general optimization problems that arise in a great deal of contexts. Furthermore, we have shown a variety of modeling features that can be accommodated in the framework. Hence, the proposed methodology provides a tool for researchers and practitioners to rely on a more accurate representation of the demand when planning and designing for their systems.

5.2 Future research directions

The current framework can be seen as a Stackelberg problem in which the (single) operator plays the leader role and the individuals collectively play the follower role. Indeed, the leader knows how the follower will respond to any potential decision being made thanks to the behavioral assumption as provided by DCM. In the context of transportation, for instance, a single supplier of transportation services or regulating agency aims at determining an optimal operation plan, rates, and/or controls while taking into account user response (Fisk, 1984). Hence, the leader's optimization problem contains a nested optimization task that corresponds to the follower's optimization problem, known as upper-level and lower-level optimization task, respectively, which define a bilevel programming problem. In this case, the problem associated with the leader (operator) is the optimization of an aggregate performance of the system on the supply-related decisions, and the follower's problem (individuals) is to select the alternative at each scenario such that the utility is maximized.

As bilevel programming problems are hard to solve, reformulations as a one-level mathematical programming problem have been proposed. The two major reformulations that can be found in the literature are Karush-Kuhn-Tucker (KKT) reformulation and the optimal value reformulation (Dempe and Zemkoho, 2013). KKT reformulation (Dempe and Dutta, 2012), which is a mathematical programming problem with complementarity constraints (MPCC), replaces the follower's problem by its KKT conditions provided that it is convex in the lower-level variables and an appropriate constraint qualification is satisfied. This reformulation introduces new variables, and therefore the two problems are not locally equivalent. The optimal value reformulation (Outrata, 1988) is obtained by replacing the lower-level solution set by its description via the optimal value function, which leads to a reformulation completely equivalent to the initial problem. It should be noticed that the optimal value function is typically non-smooth. Specific reformulations have recently received attention (e.g., Li and Guo, 2017 for mixed integer bilevel programming problems).

We find multiples examples of DCM present in the lower-level optimization task in traffic control, where the demand model refers to the route choice behavior (e.g., Sun et al., 2006 in optimal signal control). In the methodology described in this thesis, a set of mixed-integer linear constraints characterizing the DCM is embedded in a single-level formulation associated with the operator. Nonetheless, and given the increasing interest on bilevel programming problems due to their presence in several practical applications and the potential of evolutionary algorithms tackling these problems (Sinha et al., 2017), the modeling framework could be as well investigated and extended from that perspective.

The interactions between the individuals and the supply-related decisions to be optimized are not the only interactions that exist in the system. The predominant industry structure in the real world is oligopolistic, where a certain number of operators compete for the same pool of individuals and each operator aims at optimizing its own performance function. Steps in this direction have already been taken in Bortolomiol et al. (2019). They model competition among operators with a mixed-integer optimization model based on the fixed-point iteration algorithm, and are able to identify equilibrium solutions (i.e., stationary states of the system where no operator has an incentive to change their decisions) of oligopolistic markets in two transportation case studies.

DCM are fundamentally grounded in individual behavior, and do not take into account the interdependence that exists between the individuals' choices. In reality, however, individuals are influenced, for instance, by people of similar socioeconomic status who are nearby. These influences are especially relevant in the context of transportation, as the choices of groups such as households or couples play an important role in travel demand analysis. From a modeling point of view, these social interactions can be incorporated in the utility function in the way described by Brock and Durlauf (2001). They consider that the utility associated with each individual is composed of a private utility associated with the own choice and a social utility that depends on other individuals' choices. The choice-based optimization framework allows to accommodate social influence strategies thanks to the choice variables. However, we expect that the presence of such variables in the utility functions will increase the complexity of the model, and simplifications on the way social interactions are represented might need to be made.

In addition to the interactions between the actors, it might also be relevant to account for the evolution of the system over time. The operator may change its decisions over time (e.g., an airline might adjust it prices as a function of the remaining number of seats), and the choices of individuals also depend on the evolution of the market and personal experiences in the past, and might even include speculating strategies towards the future. Dynamics are specially relevant in revenue management (RM), as both dynamic pricing and dynamic assortment customization have the potential to significantly increase revenue (Bernstein et al., 2015, Talebian et al., 2014). Hence, adding a time horizon to the framework will enable the operators to gain additional control over its decisions and will provide a more realistic representation of the decision making process of individuals. Moreover, adaptive learning strategies (Hopkins, 2007) could as well be incorporated to reevaluate the behavior of individuals as they might make use of some learning rules when they are faced to repeated decision tasks. Again, we expect the inclusion of dynamic interactions to increase the complexity of the formulation, especially if a refined time discretization is assumed. Furthermore, it might be very data demanding.

As the modeling framework considers more interactions, the necessity for heuristic algorithms to speed up the solution approach becomes more apparent. We have shown that Lagrangian decomposition provides a relevant scheme to address the tractability of the model, but other techniques could also be explored or even combined. For instance, due to the presence of integer (binary) variables in the formulation, Benders decomposition strategies could also be investigated (e.g., capacity as a decision variables, as it entails a reduced set of binary variables). Moreover, as discussed in Chapter 3, routines that parallelize the solving of the decomposed subproblems could be implemented in order to reduce the total computational time. Additionally, variance reduction methods, such as control variates or importance sampling can be carried out in order to reduce the number of simulation draws while preserving a certain degree of precision of the model.

We are convinced that many real-world applications can benefit from the procedures developed in this thesis. The choice-based optimization framework allows to simultaneously consider performance indicators of the operator and measures of social welfare and distributional analysis that assess the level of satisfaction of individuals. We highlight two relevant examples in the field of transportation: overbooking strategies in airlines and rebalancing operations in vehicle sharing systems.

An important result drawn from theoretical models and empirical research in revenue management in the context of airlines is that overselling can effectively counteract revenue losses due to passenger no-shows and late cancellations. In practice, intentionally overbooking is an important strategy for airlines to manage their perishable seats, but it comes with the challenge of balancing the possible consequences of spoilage and denied boarding. Indeed, ineffective and poorly executed overbooking situations can be costly, causing not only financial losses, but also damaging customer goodwill (Ma et al., 2019).

Vehicle sharing systems are currently available in many cities around the world, and can be easily booked by users through mobile phone applications. They provide a convenient and affordable service, and in the case of car sharing systems, they also give drivers an incentive to minimize their vehicle use and to rely on alternative travel options (Litman, 2000). Existing research has mainly focused on vehicle relocation and rebalancing strategies (Huang et al., 2018), but multiple works in the literature stress that the estimation of the spatial and temporal distribution of the demand is an essential aspect that needs to be incorporated for a better performance of the system (Kaspi et al., 2016, Boyacı et al., 2015).

In conclusion, the models and algorithms developed in this thesis open the door to a great deal of potential research that will give rise to operations research models to have access to the powerful and sophisticated models developed using discrete choice theory. We believe that this is an important step towards building a bridge between the two communities that will enhance their synergies.

A

Notations of the choice-based optimization framework

This appendix includes the main notations of the choice-based optimization framework introduced in Chapter 2. The same notations are kept in the other chapters.

Table A.1: Main notations used in the choice-based profit maximization problem (sets, parameters, variables and aggregated quantities). For the sake of simplicity, we redefine the set of alternatives $\mathcal{C}$ as the set of duplicates associated with each capacity level q, and we therefore denote c_{iq}, f_{iq} and v_{iq} simply by c_i, f_i and v_i, respectively.

Name	Description	Section
$\mathcal{C}$	Set of all potential alternatives (indexed by i, $i = 0$ denotes the opt-out)	2.2.1
N	Number of individuals in the population (indexed by $n \geq 1$)	2.2.1
$\mathcal{C}_n$	Set of the alternatives considered by individual n	2.2.1
R	Number of draws from the distribution of ε_{in} (indexed by r)	2.2.2
Q	Number of capacity levels (indexed by q)	2.2.3
J	Number of alternatives in $\mathcal{C}$	2.2.3
L_{in}	Number of binary variables characterizing p_{in} (indexed by ℓ)	2.3
ξ_{inr}	Draw from the distribution of ε_{in}	2.2.2
ℓ_{inr}	Lower bound on U_{inr}	2.2.2
m_{inr}	Upper bound on U_{inr}	2.2.2
ℓ_{nr}	Smallest lower bound across alternatives	2.2.2
m_{nr}	Largest upper bound across alternatives	2.2.2
M_{inr}	$m_{inr} - \ell_{nr}$	2.2.2
M_{nr}	$m_{nr} - \ell_{nr}$	2.2.2
c_i	Capacity of alternative i	2.2.1
k	Number of decimals associated with p_{in} (precision)	2.3
a_{in}	Lower bound on p_{in} (continuous case) or on $10^k p_{in}$ (discrete case)	2.3
b_{in}	Upper bound on p_{in} (continuous case) or on $10^k p_{in}$ (discrete case)	2.3
f_i	Fixed cost associated with alternative i	2.3
v_i	Cost per sold unit of i	2.3
U_{in}	Utility associated with alternative i by individual n	2.2.1
V_{in}	Deterministic part of the utility function U_{in}	2.2.1
ε_{in}	Error term of the utility function U_{in}	2.2.1
U_{inr}	Utility associated with alternative i by individual n in scenario r	2.2.2
y_{in}	Availability at operator level of alternative i to individual n	2.2.1, 2.2.2
y_{inr}	Availability at scenario level of alternative i to individual n in scenario r	2.2.2
z_{inr}	Discounted utility associated with alternative i of individual n in scenario r	2.2.2
w_{inr}	Choice variable associated with alternative i by individual n in scenario r	2.2.2
U_{nr}	Highest value of z_{inr}	2.2.2
p_{in}	Price that individual n has to pay to access alternative $i \in \mathcal{C}_n \setminus \{0\}$	2.3
η_{inr}	Continuous variable capturing $p_{in}w_{inr}$ (continuous case)	2.3
$\lambda_{in\ell}$	Binary variable characterizing p_{in} (discrete case)	2.3
$\alpha_{inr\ell}$	Binary variable capturing $\lambda_{in\ell}w_{inr}$	2.3
D_i	Expected demand of alternative i	2.2.1, 2.2.2
G_i	Expected gain obtained from alternative $i \in \mathcal{C} \setminus \{0\}$	2.3
C_i	Total cost associated with alternative $i \in \mathcal{C} \setminus \{0\}$	2.3

B

Convergence of the choice-based optimization framework

This appendix outlines the proof of the convergence property described in Section 2.2.5. The key idea is that the sequence of optimal solutions of the approximated optimization problems (where the demand is being approximated with the mixed-integer linear formulation proposed in Section 2.2) converges to a feasible solution of the original problem (where the demand is obtained by relying on the choice probabilities).

Definition of the problems

We consider a general optimization model defined as:

$$\min_{x_s, x_e, D} f(x_s, x_e, D) \tag{B.1}$$

$$\text{subject to } D = h(x_e), \tag{B.2}$$

$$(x_s, x_e) \in X, \tag{B.3}$$

$$(x_s, x_e, D) \in Y. \tag{B.4}$$

The demand model $D \in \mathbb{R}^{\Delta}$ is characterized in constraint (B.2) with a continuous function:

$$\begin{aligned} h : \mathbb{R}^{E} &\longrightarrow \mathbb{R}^{\Delta} \\ x_e &\longmapsto h(x_e) = D. \end{aligned} \tag{B.5}$$

Notice that h may also depend on other variables that are exogenous to the optimization problem (x_d) and, therefore, not mentioned here. The objective function f relates the decision variables to an aggregate performance of the system:

$$\begin{aligned} f : \mathbb{R}^S \times \mathbb{R}^E \times \mathbb{R}^\Delta &\longrightarrow \quad \mathbb{R}, \\ (x_s, x_e, D) &\longmapsto \quad f(x_s, x_e, D), \end{aligned} \tag{B.6}$$

and is assumed to be continuous. The constraints of the optimization problem are expressed via sets. We assume that $X \subseteq \mathbb{R}^S \times \mathbb{R}^E$ is a closed subset that represents the constraints that do not involve the demand model (integrality constraints can be characterized by this subset), and $Y \subseteq \mathbb{R}^S \times \mathbb{R}^E \times \mathbb{R}^\Delta$ is a closed and bounded subset (and consequently compact) that represents the constraints that involve the demand model.

In our particular case, we assume that the demand is represented with a discrete choice model (DCM). Hence, the function $h = (h^1, \ldots, h^J)$, where J is the number of alternatives in the choice set $\mathcal{C}$, is defined as follows:

$$\begin{aligned} h^i : \mathbb{R}^E &\longrightarrow \quad \mathbb{R} \\ x_e &\longmapsto \quad D^i = \sum_n P_n(i|x^d_{in}, x^e_{in}), \end{aligned} \tag{B.7}$$

where $P_n(i|x^d_{in}, x^e_{in})$ is the choice probability as defined in (2.3). Furthermore, we assume the optimization problem defined in Section 2.2.5. It consists of a linear objective function f, as defined by (2.32), and the feasible configurations of the variables are identified by X, which is defined by the linear constraints (2.33) and the (potential) integrality constraints on a subset of the variables x_e and x_s (2.35)–(2.36), and Y, which is the polyhedron defined by the linear constraints (2.34) and the bounds on the variables (2.1) and (2.37). This specification is motivated by the need to solve to global optimality, which requires a convex formulation, or better, linear.

We define the approximated optimization problem $\mathcal{P}_R$ as:

$$\min_{x_s, x_e, D} f(x_s, x_e, D) \tag{B.8}$$

$$\text{subject to } D = h_R(x_e), \tag{B.9}$$

$$(x_s, x_e) \in X, \tag{B.10}$$

$$(x_s, x_e, D) \in Y, \tag{B.11}$$

where the demand model is approximated with a continuous function h_R:

$$\begin{aligned} h_R : \mathbb{R}^E &\longrightarrow \quad \mathbb{R}^\Delta \\ x_e &\longmapsto \quad h_R(x_e) =: D. \end{aligned} \tag{B.12}$$

It is such that, for each $x_e \in \mathbb{R}^E$, we have

$$\lim_{R \to \infty} h_R(x_e) = h(x_e). \tag{B.13}$$

We assume that there exists $\bar{R}$ such that $\mathcal{P}_R$ is feasible $\forall R \geq \bar{R}$, and that at least an optimal solution exists. We denote it by $(x_s^R,\, x_e^R,\, D^R)$. In the following, when we refer to R we implicitly assume $R \geq \bar{R}$.

The function h_R that approximates the demand model in our case is defined as

$$\begin{aligned} h_R^i : \mathbb{R}^E &\longrightarrow \mathbb{R} \\ x_e &\longmapsto D_R^i = \frac{1}{R} \sum_n \sum_r w_{inr}, \end{aligned} \tag{B.14}$$

where w_{inr} are the binary variables that determine the choice (see Section 2.2.2). As discussed in Section 2.2.2, the choice probabilities calculated with the simulation-based linearization of the DCM converge when $R \to \infty$ to the choice probabilities of the DCM (as a consequence of the law of large numbers). Hence, h_R defined by (B.14) satisfies condition (B.13).

Notice that problems $\mathcal{P}$ and $\mathcal{P}_R$ have a different feasible set. In order for both problems to have the same feasible set, we define the following penalized versions of the problems.

For a given penalty parameter $c_k \geq 0$, the penalized version of problem $\mathcal{P}$ is denoted by $\mathcal{Q}_k$ and defined as:

$$\min_{x_s, x_e, D} q_k(x_s, x_e, D) = f(x_s, x_e, D) + c_k \|D - h(x_e)\| \tag{B.15}$$

$$\text{subject to } (x_s, x_e) \in X, \tag{B.16}$$

$$(x_s, x_e, D) \in Y. \tag{B.17}$$

Similarly, given a penalty parameter $c_k^R \geq 0$, the penalized version of problem $\mathcal{P}_R$ is denoted by $\mathcal{Q}_k^R$ and defined as:

$$\min_{x_s, x_e, D} q_k^R(x_s, x_e, D) = f(x_s, x_e, D) + c_k^R \|D - h_R(x_e)\| \tag{B.18}$$

$$\text{subject to } (x_s, x_e) \in X, \tag{B.19}$$

$$(x_s, x_e, D) \in Y. \tag{B.20}$$

The penalized problems $\mathcal{Q}_k$ and $\mathcal{Q}_k^R$ are asymptotically equivalent to the original problems $\mathcal{P}$ and $\mathcal{P}_R$, respectively, in the sense determined by the following lemma.

Lemma 1 *Consider any monotonically increasing sequence $\{c_k\}_k$ of penalty parameters $c_k \geq 0$, with $\lim_{k \to +\infty} c_k = +\infty$. Consider a sequence $\{(x_s^k, x_e^k, D^k)\}_k$ such that, for each*

k, (x_s^k, x_e^k, D^k) is an optimal solution of the penalized problem $\mathcal{Q}_k$. Then, each limit point of the sequence $\{(x_s^k, x_e^k, D^k)\}_k$ is an optimal solution of problem $\mathcal{P}$.

Proof. This result is proven in Theorem 19.2 in Bierlaire (2015). □

Notice that Lemma 1 also applies to each approximation problem $\mathcal{P}_R$. In order to synchronize the penalty parameters across all problems, we use the same sequence $\{\gamma_k\}_k$ for each problem. It is defined such that

$$\gamma_k \geq c_k, \tag{B.21}$$

$$\gamma_k \geq c_k^R, \qquad \forall R. \tag{B.22}$$

As both sequences $\{c_k\}_k$ and $\{c_k^R\}_k$ go to infinity as k goes to infinity, so does the sequence $\{\gamma_k\}_k$.

Lemma 2 *For each k, and each (x_s, x_e, D) feasible for $\mathcal{P}$, we have*

$$\lim_{R\to\infty} q_k^R(x_s, x_e, D) = f(x_s, x_e, D). \tag{B.23}$$

Proof. We just need to consider the definition of q_k^R in (B.18) and use the feasibility of (x_s, x_e, D) for $\mathcal{P}$ and (B.13). We have

$$\begin{aligned}\|q_k^R(x_s, x_e, D) - f(x_s, x_e, D)\| &= \gamma_k \|D - h_R(x_e)\| \\ &= \gamma_k \|D - h_R(x_e) + h(x_e) - h(x_e)\| \\ &\leq \gamma_k \|D - h(x_e)\| + \gamma_k \|h(x_e) - h_R(x_e)\|. \end{aligned} \tag{B.24}$$

The first term in (B.24) is zero from the feasibility of (x_s, x_e, D) for $\mathcal{P}$, and the second term in (B.24) goes to zero when $R \to \infty$, from (B.13). □

Asymptotic feasibility

Lemma 3 *Denote by $(x_s^R,\, x_e^R,\, D^R)$ the optimal solution of problem $\mathcal{P}_R$. As the sequence $\{(x_s^R, x_e^R, D^R)\}_R$ is bounded because $(x_s^R, x_e^R, D^R) \in Y, \forall R$, it contains convergent subsequences. Consider (x_s^+, x_e^+, D^+) any of the accumulation points. Then, (x_s^+, x_e^+, D^+) is feasible for $\mathcal{P}$.*

Proof. As X and Y are closed, $(x_s^+, x_e^+) \in X$ and $(x_s^+, x_e^+, D^+) \in Y$. So we just need to show that $D^+ = h(x_e^+)$. We have

$$\begin{aligned} |h_R(x_e^R) - h(x_e^+)| &= |h_R(x_e^R) - h_R(x_e^+) + h_R(x_e^+) - h(x_e^+)| \\ &\leq |h_R(x_e^R) - h_R(x_e^+)| + |h_R(x_e^+) - h(x_e^+)|. \end{aligned} \tag{B.25}$$

The term $|h_R(x_e^R) - h_R(x_e^+)|$ in (B.25) converges to 0 by continuity of h_R, and the term $|h_R(x_e^+) - h(x_e^+)|$ in (B.25) converges to 0 by (B.13). Therefore, we have

$$D^+ = \lim_{R\to\infty} D^R = \lim_{R\to\infty} h_R(x_e^R) = h(x_e^+), \tag{B.26}$$

where the second equality in (B.26) comes from the feasibility of D^R for problem $\mathcal{P}_R$. $\square$

Lemma 4 *Again, denote by* (x_s^R, x_e^R, D^R) *the optimal solution of problem* $\mathcal{P}_R$ *and by* (x_s^+, x_e^+, D^+) *an accumulation point of the sequence* $\{(x_s^R, x_e^R, D^R)\}_R$*. Consider* R *and* k*, and denote by* $(x_s^R(k), x_e^R(k), D^R(k))$ *an optimal solution of* $\mathcal{Q}_k^R$*. Similarly to* $\mathcal{P}_R$*, denote by* $(x_s^+(k), x_e^+(k), D^+(k))$ *an accumulation point of the sequence* $\{(x_s^R(k), x_e^R(k), D^R(k))\}_R$ *when* $R \to \infty$*. Then,*

$$\lim_{k\to\infty} q_k(x_s^+(k), x_e^+(k), D^+(k)) = f(x_s^+, x_e^+, D^+). \tag{B.27}$$

Proof. From the optimality of $(x_s^R(k), x_e^R(k), D^R(k))$, we have

$$q_k^R(x_s^R(k), x_e^R(k), D^R(k)) \leq q_k^R(x_s^R, x_e^R, D^R). \tag{B.28}$$

From the feasibility of (x_s^R, x_e^R, D^R) for $\mathcal{P}_R$, the right hand side of (B.28) is

$$q_k^R(x_s^R, x_e^R, D^R) = f(x_s^R, x_e^R, D^R),$$

so that

$$q_k^R(x_s^R(k), x_e^R(k), D^R(k)) \leq f(x_s^R, x_e^R, D^R). \tag{B.29}$$

Consider the left-hand side of (B.28). We have

$$q_k^R(x_s^R(k), x_e^R(k), D^R(k)) = f(x_s^R(k), x_e^R(k), D^R(k)) + \gamma_k \|D^R(k) - h_R(x_e^R(k))\|. \tag{B.30}$$

From (B.29), the sequence is bounded from above. Therefore, the second term vanishes when $k \to \infty$, and $(x_s^R(k), x_e^R(k), D^R(k))$ becomes asymptotically feasible for $\mathcal{P}_R$. Hence, from the optimality of (x_s^R, x_e^R, D^R) for $\mathcal{P}_{\mathcal{R}}$ and (B.29), we have

$$\begin{aligned} f(x_s^R, x_e^R, D^R) &\leq \lim_{k\to\infty} q_k^R(x_s^R(k), x_e^R(k), D^R(k)) \\ &\leq f(x_s^R, x_e^R, D^R). \end{aligned}$$

As a consequence, if $R \to \infty$, we obtain that

$$\lim_{k\to\infty} q_k(x_s^+(k), x_e^+(k), D^+(k)) = f(x_s^+, x_e^+, D^+). \tag{B.31}$$

□

Asymptotic optimality

Theorem 5 *Denote by (x_s^*, x_e^*, D^*) an optimal solution of problem $\mathcal{P}$ and (x_s^R, x_e^R, D^R) an optimal solution of problem $\mathcal{P}_R$. As the sequence $\{(x_s^R, x_e^R, D^R)\}_R$ is bounded because $(x_s^R, x_e^R, D^R) \in Y, \forall R$, it contains convergent subsequences. Consider (x_s^+, x_e^+, D^+) any of the accumulation points. Then, we have that (x_s^+, x_e^+, D^+) is feasible for $\mathcal{P}$ and*

$$f(x_s^+, x_e^+, D^+) = f(x_s^*, x_e^*, D^*) \tag{B.32}$$

so that (x_s^+, x_e^+, D^+) is an optimal solution of $\mathcal{P}$.

Proof. The feasibility is shown by Lemma 3. Therefore,

$$f(x_s^+, x_e^+, D^+) \geq f(x_s^*, x_e^*, D^*). \tag{B.33}$$

Consider R and k, and define $(x_s^R(k), x_e^R(k), D^R(k))$ an optimal solution of $\mathcal{Q}_k^R$. We denote by $(x_s^+(k), x_e^+(k), D^+(k))$ an accumulation point of the sequence $\{(x_s^R(k), x_e^R(k), D^R(k))\}_R$, when $R \to \infty$. From the optimality of $(x_s^R(k), x_e^R(k), D^R(k))$, we have

$$q_k^R(x_s^R(k), x_e^R(k), D^R(k)) \leq q_k^R(x_s^*, x_e^*, D^*). \tag{B.34}$$

For the left hand side of (B.34), we have

$$\begin{aligned} \lim_{R\to\infty} q_k^R(x_s^R(k), x_e^R(k), D^R(k)) = &\lim_{R\to\infty} f(x_s^R(k), x_e^R(k), D^R(k)) \\ &+ \gamma_k \lim_{R\to\infty} \|D^R(k) - h_R(x_e^R(k))\| \\ =&f(x_s^+(k), x_e^+(k), D^+(k)) \\ &+ \gamma_k \|D^+(k) - h(x_e^+(k))\| \\ =&q_k(x_s^+(k), x_e^+(k), D^+(k)). \end{aligned} \tag{B.35}$$

For the right hand side of (B.34), we use Lemma 2, so that we have

$$\lim_{R\to\infty} q_k^R(x_s^*, x_e^*, D^*) = f(x_s^*, x_e^*, D^*). \tag{B.36}$$

Thus, combining (B.35) and (B.36) we have

$$q_k(x_s^+(k), x_e^+(k), D^+(k)) \leq f(x_s^*, x_e^*, D^*). \tag{B.37}$$

From Lemma 4, when $k \to \infty$, we have

$$f(x_s^+, x_e^+, D^+) \leq f(x_s^*, x_e^*, D^*),$$

which combined with (B.33) proves the result. □

Bibliography

Ahipaşaoğlu, S. D., Arıkan, U., and Natarajan, K. (2019). Distributionally robust markovian traffic equilibrium. *Transportation Science*, 53(6):1546–1562.

Anas, A. and Lindsey, R. (2011). Reducing urban road transportation externalities: Road pricing in theory and in practice. *Review of Environmental Economics and Policy*, 5:66–88.

Andersson, S.-E. (1998). Passenger choice analysis for seat capacity control: a pilot project in scandinavian airlines. *International Transactions in Operational Research*, 5(6):471 – 486.

Ardekani, S. (1991). Economic evaluation of toll plaza operations. *Tandem*, 600(700):700–800.

Aros-Vera, F., Marianov, V., and Mitchell, J. E. (2013). p-hub approach for the optimal park-and-ride facility location problem. *European Journal of Operational Research*, 226(2):277 – 285.

Atasoy, B., Glerum, A., and Bierlaire, M. (2013). Attitudes towards mode choice in switzerland. *disP - The Planning Review*, 49(2):101–117.

Atasoy, B., Salani, M., and Bierlaire, M. (2014). An integrated airline scheduling, fleeting, and pricing model for a monopolized market. *Computer-Aided Civil and Infrastructure Engineering*, 29(2):76–90.

Baillon, J.-B. and Cominetti, R. (2008). Markovian traffic equilibrium. *Mathematical Programming*, 111(1-2):33–56.

Barahona, F. and Anbil, R. (2000). The volume algorithm: producing primal solutions with a subgradient method. *Mathematical Programming*, 87(3):385–399.

Basso, L. J., Guevara, C. A., Gschwender, A., and Fuster, M. (2011). Congestion pricing, transit subsidies and dedicated bus lanes: Efficient and practical solutions to congestion. *Transport Policy*, 18(5):676–684.

Basso, L. J. and Jara-Díaz, S. R. (2012). Integrating congestion pricing, transit subsidies and mode choice. *Transportation Research Part A: Policy and Practice*, 46(6):890 – 900.

Basso, L. J. and Silva, H. E. (2014). Efficiency and substitutability of transit subsidies and other urban transport policies. *American Economic Journal: Economic Policy*, 6(4):1–33.

Becker, F., Danaf, M., Song, X., Atasoy, B., and Ben-Akiva, M. (2018). Bayesian estimator for logit mixtures with inter- and intra-consumer heterogeneity. *Transportation Research Part B: Methodological*, 117:1 – 17.

Bellei, G., Gentile, G., and Papola, N. (2002). Network pricing optimization in multiuser and multimodal context with elastic demand. *Transportation Research Part B: Methodological*, 36(9):779 – 798.

Ben-Akiva, Moshe E., . and Lerman, S. R. (1985). *Discrete choice analysis: theory and application to travel demand / Moshe Ben-Akiva, Steven R. Lerman*. MIT Press Cambridge, Mass.

Ben-Akiva, M. and Boccara, B. (1995). Discrete choice models with latent choice sets. *International Journal of Research in Marketing*, 12(1):9 – 24. Consideration sets.

Benati, S. (1999). The maximum capture problem with heterogeneous customers. *Computers & Operations Research*, 26(14):1351 – 1367.

Benati, S. and Hansen, P. (2002). The maximum capture problem with random utilities: Problem formulation and algorithms. *European Journal of Operational Research*, 143(3):518–530.

Benders, J. F. (1962). Partitioning procedures for solving mixed-variables programming problems. *Numerische Mathematik*, 4(1):238–252.

Bernstein, F., Kök, A. G., and Xie, L. (2015). Dynamic assortment customization with limited inventories. *Manufacturing & Service Operations Management*, 17(4):538–553.

Bertsekas, D. P. (1996). *Constrained Optimization and Lagrange Multiplier Methods (Optimization and Neural Computation Series)*. Athena Scientific, 1 edition.

Bierlaire, M. (2015). *Optimization: principles and algorithms*. EPFL Press.

Bierlaire, M. and Lurkin, V. (2017). *Introduction to Disaggregate Demand Models*, chapter 3, pages 48–67.

Binder, S., Maknoon, Y., and Bierlaire, M. (2017). Exogenous priority rules for the capacitated passenger assignment problem. *Transportation Research Part B: Methodological*, 105:19–42.

Blanchet, J., Gallego, G., and Goyal, V. (2016). A markov chain approximation to choice modeling. *Operations Research*, 64(4):886–905.

Borger, B. D. and Wouters, S. (1998). Transport externalities and optimal pricing and supply decisions in urban transportation: a simulation analysis for belgium. *Regional Science and Urban Economics*, 28(2):163 – 197.

Bortolomiol, S., Lurkin, V., and Bierlaire, M. (2019). A disaggregate choice-based approach to find epsilon-equilibria of oligopolistic markets. Technical Report TRANSP-OR 191107, Transport and Mobility Laboratory, Ecole Polytechnique Fédérale de Lausanne.

Boyacı, B., Zografos, K. G., and Geroliminis, N. (2015). An optimization framework for the development of efficient one-way car-sharing systems. *European Journal of Operational Research*, 240(3):718 – 733.

Brock, W. A. and Durlauf, S. N. (2001). Discrete choice with social interactions. *The Review of Economic Studies*, 68(2):235.

Camerini, P. M., Fratta, L., and Maffioli, F. (1975). *On improving relaxation methods by modified gradient techniques*, pages 26–34. Springer Berlin Heidelberg, Berlin, Heidelberg.

Carøe, C. C. and Schultz, R. (1999). Dual decomposition in stochastic integer programming. *Operations Research Letters*, 24(1-2):37–45.

Carøe, C. C. and Tind, J. (1998). L-shaped decomposition of two-stage stochastic programs with integer recourse. *Mathematical Programming*, 83(1-3):451–464.

Chen, A., Pravinvongvuth, S., Xu, X., Ryu, S., and Chootinan, P. (2012). Examining the scaling effect and overlapping problem in logit-based stochastic user equilibrium models. *Transportation Research Part A: Policy and Practice*, 46(8):1343 – 1358.

Christidis, P. and Ibáñez-Rivas, J. N. (2012). Measuring road congestion. Technical report, European Union. Institute for Prospective and Technological Studies, Joint Research Centre.

Cohen, G. and Zhu, D. L. (1984). Decomposition coordination methods in large scale optimization problems. the nondifferentiable case and the use of augmented lagrangians. *Advances in large scale systems*, (1):203–266.

Crainic, T. G., Hewitt, M., and Rei, W. (2014). Scenario grouping in a progressive hedging-based meta-heuristic for stochastic network design. *Computers & Operations Research*, 43:90 – 99.

Crainic, T. G., Rei, W., Hewitt, M., and Maggioni, F. (2016). Partial benders decomposition strategies for two-stage stochastic integer programs. Technical Report CIRRELT-2016-37, Centre interuniversitaire de recherche sur les réseaux d'entreprise, la logistique et le transport (CIRRELT).

Damla Ahipaşaoğlu, S., Arıkan, U., and Natarajan, K. (2016). On the flexibility of using marginal distribution choice models in traffic equilibrium. *Transportation Research Part B: Methodological*, 91:130 – 158.

Dantzig, G. B. and Wolfe, P. (1960). Decomposition principle for linear programs. *Operations Research*, 8(1):101–111.

De Borger, B. and Swysen, D. (1999). Public transport subsidies versus road pricing: An empirical analysis for interregional transport in belgium. *International Journal of Transport Economics*, 26(1):55–89.

de Palma, A., Kilani, M., and Lindsey, R. (2005). Congestion pricing on a road network: A study using the dynamic equilibrium simulator metropolis. *Transportation Research Part A: Policy and Practice*, 39(7):588 – 611.

de Palma, A., Proost, S., Seshadri, R., and Ben-Akiva, M. (2018). Congestion tolling - dollars versus tokens: A comparative analysis. *Transportation Research Part B: Methodological*, 108:261 – 280.

Dempe, S. and Dutta, J. (2012). Is bilevel programming a special case of a mathematical program with complementarity constraints? *Mathematical programming*, 131(1-2):37–48.

Dempe, S. and Zemkoho, A. B. (2013). The bilevel programming problem: reformulations, constraint qualifications and optimality conditions. *Mathematical Programming*, 138(1-2):447–473.

Duran, M. A. and Grossmann, I. E. (1986). An outer-approximation algorithm for a class of mixed-integer nonlinear programs. *Mathematical Programming*, 36(3):307–339.

Erlenkotter, D. (1978). A dual-based procedure for uncapacitated facility location. *Operations Research*, 26(6):992–1009.

Escudero, L. F., Garín, M. A., Pérez, G., and Unzueta, A. (2013). Scenario cluster decomposition of the lagrangian dual in two-stage stochastic mixed 0–1 optimization. *Computers & Operations Research*, 40(1):362–377.

Escudero, L. F., Garín, M. A., and Unzueta, A. (2016). Cluster lagrangean decomposition in multistage stochastic optimization. *Computers & Operations Research*, 67:48 – 62.

Farooq, B., Bierlaire, M., Hurtubia, R., and Flötteröd, G. (2013). Simulation based population synthesis. *Transportation Research Part B: Methodological*, 58(C):243–263.

Federal Highway (1980). *The Urban transportation planning system (UTPS) : an introduction for management.*

Feldman, J. B. and Topaloglu, H. (2017). Revenue management under the markov chain choice model. *Operations Research*, 65(5):1322–1342.

Fernández-Antolín, A., Lurkin, V., de Lapparent, M., and Bierlaire, M. (2017). Discrete-continuous maximum likelihood for the estimation of nested logit models. In *Proceedings of the 16th Swiss Transport Research Conference*, Ascona, Switzerland.

Fischetti, M. and Lodi, A. (2010). Heuristics in mixed integer programming. *Wiley Encyclopedia of Operations Research and Management Science*.

Fisher, M. L. (1973). Optimal solution of scheduling problems using lagrange multipliers: Part i. *Operations Research*, 21(5):1114–1127.

Fisher, M. L. (1981a). The lagrangian relaxation method for solving integer programming problems. *Management Science*, 27(1):1–18.

Fisher, M. L. (1981b). The lagrangian relaxation method for solving integer programming problems. *Management science*, 27(1):1–18.

Fisher, M. L., Jaikumar, R., and Van Wassenhove, L. N. (1986). A multiplier adjustment method for the generalized assignment problem. *Management Science*, 32(9):1095–1103.

Fisher, M. L., Northup, W. D., and Shapiro, J. F. (1975). *Using duality to solve discrete optimization problems: Theory and computational experience*, pages 56–94. Springer Berlin Heidelberg, Berlin, Heidelberg.

Fisk, C. (1984). Game theory and transportation systems modelling. *Transportation Research Part B: Methodological*, 18(4-5):301–313.

Fosgerau, M., McFadden, D., and Bierlaire, M. (2013). Choice probability generating functions. *Journal of Choice Modelling*, 8:1 – 18.

Frangioni, A., Gendron, B., and Gorgone, E. (2017). On the computational efficiency of subgradient methods: a case study with lagrangian bounds. *Mathematical Programming Computation*, 9(4):573–604.

Freire, A. S., Moreno, E., and Yushimito, W. F. (2016). A branch-and-bound algorithm for the maximum capture problem with random utilities. *European Journal of Operational Research*, 252(1):204 – 212.

Gendron, B., Scutellà, M. G., Garroppo, R. G., Nencioni, G., and Tavanti, L. (2016). A branch-and-benders-cut method for nonlinear power design in green wireless local area networks. *European Journal of Operational Research*, 255(1):151–162.

Geoffrion, A. M. (1971). Duality in nonlinear programming. *SIAM Rev.*, 13(1):1–37.

Geoffrion, A. M. (1974). Lagrangean relaxation for integer programming. *Math. Programming Stud.*, (2):82–114. Approaches to integer programming.

Gilbert, F., Marcotte, P., and Savard, G. (2014a). Logit network pricing. *Computers & Operations Research*, 41:291–298.

Gilbert, F., Marcotte, P., and Savard, G. (2014b). Mixed-logit network pricing. *Computational Optimization and Applications*, 57(1):105–127.

Gilbert, F., Marcotte, P., and Savard, G. (2015). A numerical study of the logit network pricing problem. *Transportation Science*, 49(3):706–719.

Goffin, J.-L. and Vial, J.-P. (2002). Convex nondifferentiable optimization: A survey focused on the analytic center cutting plane method. *Optimization Methods and Software*, 17(5):805–867.

Gosavi, A. et al. (2015). *Simulation-based optimization*. Springer.

Guignard, M. (2003). Lagrangean relaxation. *Top*, 11(2):151–200.

Guignard, M. and Kim, S. (1987). Lagrangean decomposition: A model yielding stronger lagrangean bounds. *Math. Program.*, 39(2):215–228.

Guillaume-Gentil, S., Cazorla, M., and Mousset, A. (2015). Comptage periodiques de trafic ti - tc 2014.

Gupta, S., Seshadri, R., Atasoy, B., Prakash, A. A., Pereira, F., Tan, G., and Ben-Akiva, M. (2020). Real-time predictive control strategy optimization. *Transportation research record*, 2674(3):1–11.

Haase, K. (2009). Discrete location planning. institute of transport and logistics studies working paper. Technical report, ITLS-WP-09-07.

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

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

Hamdouch, Y., Florian, M., Hearn, D. W., and Lawphongpanich, S. (2007). Congestion pricing for multi-modal transportation systems. *Transportation Research Part B: Methodological*, 41(3):275 – 291.

Haneveld, W. K. K. and van der Vlerk, M. H. (1999). Stochastic integer programming: General models and algorithms. *Annals of Operations Research*, 85:39.

Held, M. and Karp, R. M. (1971a). The traveling-salesman problem and minimum spanning trees: Part ii. *Math. Program.*, 1(1):6–25.

Held, M. and Karp, R. M. (1971b). The traveling-salesman problem and minimum spanning trees: Part ii. *Mathematical Programming*, 1(1):6–25.

Held, M., Wolfe, P., and Crowder, H. P. (1974). Validation of subgradient optimization. *Mathematical Programming*, 6(1):62–88.

Hess, S. and Rose, J. M. (2009). Allowing for intra-respondent variations in coefficients estimated on repeated choice data. *Transportation Research Part B: Methodological*, 43(6):708 – 719.

Hopkins, E. (2007). Adaptive learning models of consumer behavior. *Journal of economic behavior & organization*, 64(3-4):348–368.

Huang, H.-J. (2000). Fares and tolls in a competitive system with transit and highway: the case with two groups of commuters. *Transportation Research Part E: Logistics and Transportation Review*, 36(4):267 – 284.

Huang, H.-J. (2002). Pricing and logit-based mode choice models of a transit and highway system with elastic demand. *European Journal of Operational Research*, 140(3):562 – 570.

Huang, K., de Almeida Correia, G. H., and An, K. (2018). Solving the station-based one-way carsharing network planning problem with relocations and non-linear demand. *Transportation Research Part C: Emerging Technologies*, 90:1 – 17.

Ibeas, A., dell'Olio, L., Bordagaray, M., and Ortúzar, J. (2014). Modelling parking choices considering user heterogeneity. *Transportation Research Part A: Policy and Practice*, 70(C):41 – 49.

Ieromonachou, P., Potter, S., and Warren, J. (2007). Analysing road pricing implementation processes in the uk and norway. *European Journal of Transport and Infrastructure Research*, 7(1):15–38. This paper previously appeared in The European Journal of Transport and Infrastructure Research.

Ingvardson, J. B., Nielsen, O. A., Raveau, S., and Nielsen, B. F. (2018). Passenger arrival and waiting time distributions dependent on train service frequency and station characteristics: A smart card data analysis. *Transportation Research Part C: Emerging Technologies*, 90:292–306.

Kaspi, M., Raviv, T., Tzur, M., and Galili, H. (2016). Regulating vehicle sharing systems through parking reservation policies: Analysis and performance bounds. *European Journal of Operational Research*, 251(3):969 – 987.

Kelley, J. E. (1960). The cutting-plane method for solving convex programs. *Journal of the Society for Industrial and Applied Mathematics*, 8(4):703–712.

Kidokoro, Y. (2010). Revenue recycling within transport networks. *Journal of Urban Economics*, 68(1):46–55.

Knight, F. H. (1924). Some fallacies in the interpretation of social cost. *The Quarterly Journal of Economics*, 38(4):582–606.

Korfmann, F. (2018). *Essays on Advanced Discrete Choice Applications.* PhD thesis, Universität Hamburg.

KPMG (2015). Kpmg toll benchmarking study 2015.

Kraus, M. (1989). The welfare gains from pricing road congestion using automatic vehicle identification and on-vehicle meters. *Journal of Urban Economics*, 25(3):261 – 281.

Labbé, M., Marcotte, P., and Savard, G. (1998). A bilevel model of taxation and its application to optimal highway pricing. *Management Science*, 44(12-part-1):1608–1622.

Laporte, G. and Louveaux, F. V. (1993). The integer l-shaped method for stochastic integer programs with complete recourse. *Operations Research Letters*, 13(3):133 – 142.

Laporte, G., Nickel, S., and Saldanha da Gama, F. (2015). *Location Science.* Springer.

Lemaréchal, C. (1989). Chapter vii nondifferentiable optimization. In *Optimization*, volume 1 of *Handbooks in Operations Research and Management Science*, pages 529 – 572. Elsevier.

Levinson, D. (2010). Equity effects of road pricing: A review. *Transport Reviews*, 30(1):33–57.

Li, C. and Guo, L. (2017). A single-level reformulation of mixed integer bilevel programming problems. *Operations Research Letters*, 45(1):1–5.

Litman, T. (2000). Evaluating carsharing benefits. *Transportation Research Record*, 1702:31–35.

Liu, Q. and Van Ryzin, G. J. (2008). On the choice-based linear programming model for network revenue management. *Manufacturing & Service Operations Management*, 10(2):288–310.

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

Lyons, G., Dudley, G., Slater, E., and Parkhurst, G. (2004). Evidence based review: Attitudes to road pricing. *Project Report. Department for Transport.*

Ma, J., Tse, Y. K., Wang, X., and Zhang, M. (2019). Examining customer perception and behaviour through social media research - an empirical study of the united airlines overbooking crisis. *Transportation Research Part E: Logistics and Transportation Review*, 127:192 – 205.

Manski, C. F. (1977). The structure of random utility models. *Theory and decision*, 8(3):229.

Marsten, R. E., Hogan, W. W., and Blankenship, J. W. (1975). The boxstep method for large-scale optimization. *Operations Research*, 23(3):389–405.

McFadden, D. (1974). Conditional Logit Analysis of Qualitative Choice Behavior. In *Frontiers in Econometrics*, pages 105–142. Academic Press, New York.

McFadden, D. (1978). Modeling the choice of residential location. *Transportation Research Record*, (673).

McFadden, D. and Train, K. (2000). Mixed mnl models for discrete response. *Journal of applied Econometrics*, 15(5):447–470.

McGill, J. I. and van Ryzin, G. J. (1999). Revenue management: Research overview and prospects. *Transportation Science*, 33(2):233–256.

Mohring, H. (1979). The benefits of reserved bus lanes, mass transit subsidies, and marginal cost pricing in alleviating traffic congestion. *Current issues in urban economics*, pages 165–195.

MOVE, D. (2014). Update of the handbook on external costs of transport. *Report for the*.

Mukoyama, T. (2013). Understanding the welfare effects of unemployment insurance policy in general equilibrium. *Journal of Macroeconomics*, 38:347 – 368.

Müller, S., Haase, K., and Kless, S. (2009). A multiperiod school location planning approach with free school choice. *Environment and Planning A*, 41(12):2929–2945.

Mulvey, J. M. and Ruszczyński, A. (1995). A new scenario decomposition method for large-scale stochastic optimization. *Operations research*, 43(3):477–490.

Nguyen, A.-T., Reiter, S., and Rigo, P. (2014). A review on simulation-based optimization methods applied to building performance analysis. *Applied Energy*, 113:1043 – 1058.

Osorio, C. and Atasoy, B. (forthcoming). Efficient simulation-based toll optimization for large-scale networks. *Transportation Science*.

Outrata, J. V. (1988). A note on the usage of nondifferentiable exact penalties in some special optimization problems. *Kybernetika*, 24(4):251–258.

Pacheco, M., Lurkin, V., Gendron, B., Sharif Azadeh, S., and Bierlaire, M. (2018). Lagrangian relaxation for the demand-based benefit maximization problem. In *18th Swiss Transport Research Conference (STRC), Ascona, Switzerland*.

Pacheco, M., Sharif Azadeh, S., and Bierlaire, M. (2016). A new mathematical representation of demand using choice-based optimization method. In *16th Swiss Transport Research Conference*.

Pacheco, M., Sharif Azadeh, S., and Bierlaire, M. (2019). Passenger satisfaction maximization within a demand-based optimization framework. In *19th Swiss Transport Research Conference (STRC), Ascona, Switzerland*.

Pacheco, M., Sharif Azadeh, S., Bierlaire, M., and Gendron, B. (2017a). Integrating advanced demand models within the framework of mixed integer linear problems: a lagrangian relaxation method for the uncapacitated case. In *17th Swiss transport research conference (STRC), Ascona, Switzerland*.

Pacheco, M., Sharif Azadeh, S., Bierlaire, M., and Gendron, B. (2017b). Integrating advanced discrete choice models in mixed integer linear optimization. Technical Report TRANSP-OR 170714, Transport and Mobility Laboratory, ENAC, EPFL.

Parry, I. W. H. and Bento, A. (2001). Revenue recycling and the welfare effects of road pricing. *The Scandinavian Journal of Economics*, 103(4):645–671.

Patriksson, M. (2015). *The traffic assignment problem: models and methods*. Courier Dover Publications.

Pel, A. J., Bliemer, M. C., and Hoogendoorn, S. P. (2009). Hybrid route choice modeling in dynamic traffic assignment. *Transportation Research Record*, 2091(1):100–107.

Pigou, A. (1920). *The Economics of Welfare*. Number v. 1 in The Economics of Welfare. Macmillan and Company, Limited.

Qian, Z. S. and Zhang, H. M. (2013). A hybrid route choice model for dynamic traffic assignment. *Networks and Spatial Economics*, 13(2):183–203.

Rahmaniani, R., Crainic, T. G., Gendreau, M., and Rei, W. (2017). The benders decomposition algorithm: A literature review. *European Journal of Operational Research*, 259(3):801 – 817.

Ratliff, R. M., Rao, B. V., Narayan, C. P., and Yellepeddi, K. (2008). A multi-flight recapture heuristic for estimating unconstrained demand from airline bookings. *Journal of Revenue and Pricing Management*, 7(2):153–171.

Reinoso, H. and Maculan, N. (1992). Lagrangean decomposition in integer linear programming: A new scheme. *INFOR: Information Systems and Operational Research*, 30(1):1–5.

Rockafellar, R. T. and Wets, R. J.-B. (1991). Scenarios and policy aggregation in optimization under uncertainty. *Mathematics of operations research*, 16(1):119–147.

Rosa, C. H. and Ruszczyński, A. (1996). On augmented lagrangian decomposition methods for multistage stochastic programs. *Annals of Operations Research*, 64(1):289–309.

Rouwendal, J. and Verhoef, E. T. (2006). Basic economic principles of road pricing: From theory to applications. *Transport Policy*, 13(2):106 – 114. Modelling of Urban Road Pricing and its Implementation.

Ruszczyński, A. and Shapiro, A. (2003). Stochastic programming models. *Handbooks in operations research and management science*, 10:1–64.

Ruszczyński, A. (2003). Decomposition methods. In *Stochastic Programming*, volume 10 of *Handbooks in Operations Research and Management Science*, pages 141 – 211. Elsevier.

Schade, J. and Schlag, B. (2003). Acceptability of urban transport pricing strategies. *Transportation Research Part F: Traffic Psychology and Behaviour*, 6(1):45 – 61.

Schön, C. (2007). Market-oriented airline service design. In *Operations Research Proceedings 2006*, pages 361–366. Springer.

Schön, C. (2008). Integrated airline schedule design, fleet assignment and pricing. *DSOR-Beiträge zur Wirtschaftsinformatik= DSOR Contributions to Information Systems*, 5:73–88.

Sheffi, Y. (1984). *Urban Transportation Networks: Equilibrium Analysis With Mathematical Programming Methods.*

Shen, Z.-J. M. and Su, X. (2007). Customer behavior modeling in revenue management and auctions: A review and new research opportunities. *Production and operations management*, 16(6):713–728.

Shepherd, S., Zhang, X., Emberger, G., Hudson, M., May, A., and Paulley, N. (2006). Designing optimal urban transport strategies: The role of individual policy instruments and the impact of financial constraints. *Transport Policy*, 13(1):49 – 65.

Sinha, A., Malo, P., and Deb, K. (2017). A review on bilevel optimization: from classical to evolutionary approaches and applications. *IEEE Transactions on Evolutionary Computation*, 22(2):276–295.

Small, K. A. (1983a). Bus priority and congestion pricing on urban expressways. *Research in Transportation Economics*, 1.

Small, K. A. (1983b). The incidence of congestion tolls on urban highways. *Journal of urban economics*, 13(1):90–111.

Smith, J. C. and Taskin, Z. C. (2008). A tutorial guide to mixed-integer programming models and solution techniques. *Optimization in Medicine and Biology*, pages 521–548.

Sun, D., Benekohal, R. F., and Waller, S. T. (2006). Bi-level programming formulation and heuristic solution approach for dynamic traffic signal optimization. *Computer-Aided Civil and Infrastructure Engineering*, 21(5):321–333.

Talebian, M., Boland, N., and Savelsbergh, M. (2014). Pricing to accelerate demand learning in dynamic assortment planning for perishable products. *European Journal of Operational Research*, 237(2):555–565.

Talluri, K. T. and Van Ryzin, G. J. (2004). Revenue management under a general discrete choice model of consumer behavior. *Management Science*, 50(1):15–33.

Talluri, K. T. and Van Ryzin, G. J. (2006). *The Theory and Practice of Revenue Management*. International Series in Operations Research & Management Science. Springer US.

TRB, H. C. M. (2010). Hcm 2010. *Transportation Research Board–special report*, 209.

Tsekeris, T. and Voß, S. (2009). Design and evaluation of road pricing: state-of-the-art and methodological advances. *NETNOMICS: Economic Research and Electronic Networking*, 10(1):5–52.

Tsekeris, T. and Voss, S. (2010). Public transport and road pricing: A survey and simulation experiments. *Public Transport*, 2:87–109.

van Ryzin, G. J. (2005). Future of revenue management: models of demand. *Journal of Revenue and Pricing Management*, 4(2):204–210.

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

Vanderbeck, F. (2000). On dantzig-wolfe decomposition in integer programming and ways to perform branching in a branch-and-price algorithm. *Operations Research*, 48(1):111–128.

Vielma, J. P. (2015). Mixed integer linear programming formulation techniques. *Siam Review*, 57(1):3–57.

Vulcano, G. J., van Ryzin, G. J., and Chaar, W. (2010). Om practice - choice-based revenue management: An empirical study of estimation and optimization. *Manufacturing & Service Operations Management*, 12(3):371–392.

Walker, J. L. (2001). *Extended discrete choice models: integrated framework, flexible error structures, and latent variables*. PhD thesis, Massachusetts Institute of Technology.

Williams, H. C. (1977). On the formation of travel demand models and economic evaluation measures of user benefit. *Environment and planning A*, 9(3):285–344.

Wollmer, R. D. (1980). Two stage linear programming under uncertainty with 0–1 integer first stage variables. *Mathematical programming*, 19(1):279–288.

Wu, D., Yin, Y., Lawphongpanich, S., and Yang, H. (2012). Design of more equitable congestion pricing and tradable credit schemes for multimodal transportation networks. *Transportation Research Part B: Methodological*, 46(9):1273 – 1287.

Zhang, Y., Berman, O., and Verter, V. (2012). The impact of client choice on preventive healthcare facility network design. *OR spectrum*, 34(2):349–370.

Meritxell Pacheco Paneque

Transport and Mobility Laboratory (TRANSP-OR)
School of Architecture, Civil and Environmental Engineering, Station 18
École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland

+41 21 693 81 00 • meritxell.pacheco@epfl.ch

Education

École Polytechnique Fédérale de Lausanne — **Lausanne**
Doctoral degree (PhD) — *Sep 2015-present*
A general framework for the integration of complex choice models into mixed integer optimization

Universitat Politècnica de Catalunya — **Barcelona**
MSc in Logistics, Transportation and Mobility — *2013-2015*
GPA = 9.1/10

Universitat Politècnica de Catalunya — **Barcelona**
BSc and MSc in Mathematics (5-year degree) — *2007-2013*
GPA = 7.0/10

Universität Mannheim — **Mannheim**
Financial mathematics — *2010-2011*
One-year exchange program

Professional experience

inLab (Universitat Politècnica de Catalunya) — **Barcelona**
Data Scientist Internship — *2014-2015*
Traffic data analysis

Konrad-Zuse-Zentrum für Informationnstechnik Berlin (ZIB) — **Berlin**
Graduate-Level Research in Industrial Projects (G-RIPS) Summer Internship — *2014*
Schedule optimization for high-speed rail in Germany (Deutsche Bahn)

Universitat Politècnica de Catalunya — **Barcelona**
International Relations Office Internship — *2013-2014*

Deloitte Consulting — **Barcelona**
Junior Consultant — *2013*
Marketing and sales

La Magnética — **Barcelona**
Junior Consultant — *2012-2013*
Data analysis for marketing

Languages

Spanish: Mother tongue

Catalan: Mother tongue

English: Independent User - Vantage (B2) — *TOEIC Certificate with a total score of 805/990*

German: Independent User - Vantage (B2) — *Escuela Oficial de Idiomas (Spanish official school of languages)*

French: Independent User (B1-B2) — *EPFL Centre de Langues (EPFL Language Centre)*

Italian: Beginner (A1-A2) — *Escuela Oficial de Idiomas (Spanish official school of languages)*

Computer skills

OS: Windows, Mac
Office: Word, Excel, Power Point
Mathematics: Matlab
Programming: C++
Statistics: R
Others: LaTeX, pandasbiogeme

Research projects

Incorporating advanced behavioral models in mixed integer linear optimization **Lausanne**
Swiss National Science Foundation *2016-2019*
This project proposes a modeling framework for the integration of advanced discrete choice models in mixed-integer linear formulations.

Teaching

Mathematical Modeling of Behavior
EPFL, Master course *2016-2020*

Discrete Choice Analysis: Predicting Individual Behavior and Market Demand
EPFL, Postgraduate one-week program *2016-2020*

Decision-aid methodologies in transportation
EPFL, Master course *2016-2017*

Introduction à l'optimisation et la recherche opérationnelle
EPFL, Bachelor course *2015-2017*

Student project supervision

Master theses

- *Disruption-caused railway timetable rescheduling problem and its solution (2019).* Oliver Buschor (EPFL).
- *Investigations on a pilot survey for modelling mobility behaviours in Switzerland (2020).* René Lugrin (EPFL).

Semester projects

- *Passenger satisfaction maximization under budget constraints (2019).* Tatiana Moavensadeh-Ghasnavi (EPFL).
- *Locating charging station for electric taxis (2019).* Oliver Buschor, Younes Bensaid (EPFL).
- *Optimization of school accessibility in developing countries (2018).* Oliver Buschor, Yassine El Ouazzani (EPFL).
- *Design of a stated-preferences survey for a high-speed vacuum transportation mode (2018).* Thibaut Richard, Martí Montesinos Ferrer (EPFL).
- *Integrating demand and supply in the context of airlines (2018).* Thibaut Richard, Gabriel Curis (EPFL).
- *Investigating the role of attitudes in the purchase of new cars (2017).* Nicola Ortelli (EPFL).
- *Modeling purchases of new cars for 2015: a comparison between countries (2016).* Martí Montesinos Ferrer (EPFL).
- *Pricing and capacity allocation strategies for a demand-based revenues maximization problem (2016).* Jonathan Lachkar (EPFL).

Publications

Technical reports

- Pacheco, M., Sharif Azadeh, S., Bierlaire, M., and Gendron, B. (2017b). Integrating advanced discrete choice models in mixed integer linear optimization. Technical Report TRANSP-OR 170714, Transport and Mobility Laboratory, ENAC, EPFL.

Papers in conference proceedings

- Pacheco, M., Sharif Azadeh, S., and Bierlaire, M. (2019). Passenger satisfaction maximization within a demand-based optimization framework. Proceedings of the 19th Swiss Transport Research Conference.
- Pacheco, M., Lurkin, V., Gendron, B., Sharif Azadeh, S., and Bierlaire, M. (2018). Lagrangian relaxation for the demand-based benefit maximization problem. Proceedings of the 18th Swiss Transport Research Conference.
- Pacheco, M., Sharif Azadeh, S., Bierlaire, M., and Gendron, B. (2017) Integrating advanced demand models within the framework of mixed integer linear problems: a Lagrangian relaxation method for the uncapacitated case. Proceedings of the 17th Swiss Transport Research Conference.
- Pacheco, M., Sharif Azadeh, S., and Bierlaire, M. (2016). A new mathematical representation of demand using choice-based optimization method. Proceedings of the 16th Swiss Transport Research Conference.

Presentations at international conferences and invited seminars

- Pacheco, M., Sharif Azadeh, S., and Bierlaire, M., Passenger satisfaction maximization within a demand-based optimization framework. 19th Swiss Transport Research Conference (STRC), May 16, 2019, Monte Verità, Ascona, Switzerland
- Pacheco, M., Gendron, B., Lurkin, V., Bierlaire, M., and Sharif Azadeh, S., A Lagrangian relaxation technique for the demand-based benefit maximization problem. Annual international conference of the German Operations Research Society (OR2018), Université libre de Bruxelles, September 14, 2018, Brussels, Belgium
- Pacheco, M., Gendron, B., Lurkin, V., Bierlaire, M., and Sharif Azadeh, S., A Lagrangian relaxation technique for the demand-based benefit maximization problem. hEART 2018, 7th Symposium of the European Association for Research in Transportation, National Technical University of Athens, September 06, 2018, Athens, Greece
- Pacheco, M., Sharif Azadeh, S., Gendron, B., Bierlaire, M., and Lurkin, V., Integrating discrete choice models in mixed integer linear programming to capture the interactions between supply and demand. 15th International Conference on Travel Behavior Research, University of California Santa Barbara, July 17, 2018, Santa Barbara, CA, USA
- Pacheco, M., Gendron, B., Lurkin, V., Sharif Azadeh, S., and Gendron, B., Lagrangian relaxation for the demand-based benefit maximization problem. Workshop on Discrete Choice Models 2018, EPFL, June 22, 2018, Lausanne, Switzerland
- Pacheco, M., Gendron, B., Lurkin, V., Sharif Azadeh, S., and Bierlaire, M., Lagrangian relaxation for the demand-based benefit maximization problem. 16th Swiss Operations Research Days, Universität Bern, June 12, 2018, Bern, Switzerland
- Pacheco, M., Gendron, B., Lurkin, V., Sharif Azadeh, S., and Bierlaire, M., A Lagrangian relaxation technique for the demand-based benefit maximization problem. 18th Swiss Transport Research Conference (STRC), May 17, 2018, Monte Verità, Ascona, Switzerland
- Pacheco, M., Sharif Azadeh, S., Bierlaire, M., and Gendron, B., Integrating advanced discrete choice models in mixed integer linear optimization. ORBEL 32 Conference , HEC Liège, February 02, 2018, Liège, Belgium
- Pacheco, M., Sharif Azadeh, S., Bierlaire, M., and Gendron, B., A Lagrangian relaxation method for a choice-based optimization framework capturing the demand-supply interaction. Annual international conference of the German Operations Research Society (OR2017), Freie Universität Berlin, September 07, 2017, Berlin, Germany
- Pacheco, M., Sharif Azadeh, S., Bierlaire, M., and Gendron, B., Integrating supply and demand within the framework of mixed integer linear problems. 21st Conference of the International Federation of Operational Research Societies, IFORS, July 19, 2017, Québec City, Canada
- Pacheco, M., Sharif Azadeh, S., Bierlaire, M., and Gendron, B., Lagrangian relaxation for the integration of discrete choice models in mixed integer linear problems. 15th Swiss Operations Research Days, Université de Fribourg, June 30, 2017, Fribourg, Switzerland

- Pacheco, M., Sharif Azadeh, S., Bierlaire, M., and Gendron, B., Integrating advanced discrete choice models in mixed integer linear optimization. Workshop on Discrete Choice Models 2017, EPFL, June 22, 2017, Lausanne, Switzerland
- Pacheco, M., Sharif Azadeh, S., Bierlaire, M., and Gendron, B., Integrating advanced demand models within the framework of mixed integer linear problems: A Lagrangian relaxation method for the uncapacitated case. 17th Swiss Transport Research Conference (STRC), May 18, 2017, Monte Verità, Ascona, Switzerland
- Pacheco, M., Sharif Azadeh, S., and Bierlaire, M., A new mathematical formulation to integrate supply and demand within a choice-based optimization framework. Seminario en el Departamento de Ingeniería de Transporte y Logística, Pontificia Universidad Católica de Chile, October 13, 2016, Santiago de Chile, Chile
- Pacheco, M., Sharif Azadeh, S., and Bierlaire, M., A new mathematical formulation to integrate supply and demand within a choice-based optimization framework. Seminario ISCI, Universidad de Chile, October 12, 2016, Santiago de Chile, Chile
- Pacheco, M., Sharif Azadeh, S., and Bierlaire, M., Integrating supply and demand within the framework of mixed integer optimization problems mixed integer optimization problems. hEART 2016, 5th Symposium of the European Association for Research in Transportation, Delft University of Technology, September 15, 2016, Delft, Netherlands
- Pacheco, M., Sharif Azadeh, S., and Bierlaire, M., A new mathematical formulation to integrate supply and demand within a choice-based optimization framework. EURO 2016, July 06, 2016, Poznan, Poland
- Pacheco, M., Sharif Azadeh, S., and Bierlaire, M., Incorporating advanced behavioral models in mixed linear optimization. Ninth Triennial Symposium on Transportation Analysis (TRISTAN IX), June 14, 2016, Oranjestad, Aruba
- Pacheco, M., Sharif Azadeh, S., and Bierlaire, M., A new mathematical formulation to integrate supply and demand within a choice-based optimization framework. 16th Swiss Transport Research Conference (STRC), May 18, 2016, Monte Verità, Ascona, Switzerland

Awards and distinctions

- EDCE PhD mobility award 2017, August 09, 2017 (EPFL).
- V Abertis International Research Award, October 14, 2016 (International Abertis Chairs Network)
- XIII Abertis Prize for Management of Transportation Infrastructures, May 24, 2016 (Abertis-UPC Chair)