UNIVERSITY OF BERGAMO

Department of Management, Information and Production Engineering

Doctoral Degree in Technology Innovation and Management
XXXIII Cycle

Modeling of Supply-Demand Interactions in the Optimization of Air Transport Networks

Thesis Supervisor
Prof. Paolo Malighetti

Author
Sebastian Birolini
PhD Candidate ID: 1024761

Academic year 2019/2020

Acknowledgements

This dissertation is the result of three intense years of studying, learning, and traveling, and it would have never been possible without the many people who, in differed ways, have supported me in this journey.

First and foremost, I would like to thank the ITSM team at the University of Bergamo. First, my supervisor, *Paolo Malighetti*, for his guidance and support. I thank Paolo for believing in me since my bachelor's thesis and for introducing me to doing research in the field of air transportation. I also thank *Mattia Cattaneo*, who has always been available to promptly and enthusiastically help with everything; from early brainstorming on new research topics, to the formal analysis, up to the drafting and writing up of papers. Besides all of that, I am most thankful to Mattia for the friendship and empathy, which delighted the late hours in the office and hours-long calls when I was abroad. I am greatly indebted to *Renato Redondi* for sharing his great knowledge on airline networks and connectivity, which has fasten my learning process and helped me identify promising research questions. I also thank Renato for his guidance and exemplary role in teaching. His dedication and commitment to students are admirable and I feel really honored to have been his student first and teaching assistant later. I would like to extend my sincere thanks to *Stefano Paleari*. Stefano's experience and careful consideration of policy implications pushed me to pursue impact through research. Moreover, I thank Stefano for getting me involved in several projects that gave me the chance to couple theoretical knowledge with practical experience.

A fundamental part of my Ph.D. was the visiting period at CITTA— Research Center on Territory Transport and Environment—in Coimbra. This period contributed to develop the ability to carry out independent research and build up my own network of collaborations. I am extremely grateful to *António Pais Antunes* for welcoming there and making me part of his research group. António was extremely supportive and an excellent mentor. Besides his many commitments, he always found time to discuss research developments and ways to address research challenges—I do remember the lunches spent talking about research (and great Portuguese food). Antonio's passion and dedication to research are truly inspiring and will accompany me as a source of inspiration in my academic career.

I am deeply grateful to *Alexandre Jacquillat* for his invaluable support. Alex's insightful feedback and challenging comments pushed me to sharpen my thinking and brought my work to a higher level. He stimulated me to look at methodological developments that I would never have believed I could address. I feel fortunate to have had the opportunity to work with Alex and I do hope to continue our collaboration in the future.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Since its inception, the air transport industry has been growing at a fast pace to finally establish itself as one of the major industries in today's global economy. The utmost relevance of air transport industry is not only due to its direct economic impact but mostly to its boosting effect on global trades and socio-economic development. Based on the most recent industry figures (IATA 2019), in 2018 airlines have accommodated about 4.4 billion origin-destination passenger journeys worldwide and the number of unique city-pairs connected by regular airline services has reached almost 22,000, corresponding to a substantial 5.4% CAGR from 2000 (1.7 billion passengers) and a 2x increase in the number of non-stop routes provided. The overall economic impact of the industry is estimated to generate more than 65.5 million jobs worldwide and US$ 1.7 trillion GDP (around 3.6 percent of the world's gross domestic product) as the sum of direct, indirect, induced, and tourism-related impacts (ATAG 2018).

Following deregulation, air transport networks have quickly grown into very complex systems involving several interacting players. Alongside the capital-intensive nature of the business and the large-scale operations, the fast-changing competitive landscape characterizing the aviation industry has made the development of effective planning schemes a highly challenging task, but more importantly, a strategic source of competitive advantage. The consolidation of air transport networks and their infrastructures has occurred in close synergy with developments in the field of operations research (OR) and quantitative methods, perhaps as in no other sector. OR has

armed practitioners with optimization-based decision support tools and contributed to significantly enhance industry performance (Barnhart et al. 2003a). Airlines in particular have long been at the forefront of innovation in operations research and, today, tactical and operational decisions are routinely supported by dedicated optimization software.

One of the main challenges of these models is to appropriately incorporate passenger demand. Demand constitutes probably the most important input to the planning processes in aviation; indeed, regardless of the sophistication of mathematical models supporting these processes, the quality of their outcome is largely, if not entirely, dependent on the accuracy of traffic estimates. Throughout the entire airline planning process, the assessment of air travel demand is pervasive and plays a vital role. Airline decision making indeed goes hand-in-hand with demand forecasting both to suitably allocate resources in the short term and to support strategic decisions on capacity planning and network developments in the long term.

The earliest mathematical models in this field have focused on capturing the complex intricacies of aviation processes to ensure feasible and realistic solutions to real-world applications (e.g., Levin 1971, Abara 1989a, Teodorović et al. 1994). These models were built under the assumption that passenger demand was fixed and given—thus treating it as a known parameter in the mathematical formulation. The consideration of inelastic demand has been taken as a (necessary) simplification for quite a long time. However, while demand surely motives and drives the provision of airline services to a large extent, on the other, the allocation of airline resources and their scheduling determine the quality of services and travel opportunities available to passengers, which in turn greatly influence their demand for air travel.

There is much evidence in the industry supporting the bi-directional relationship between air travel supply and demand. Perhaps the most evident was the advent of low-cost carriers, which, by leveraging low prices and point-to-point services, were able to stimulate demand and serve markets that were previously deemed as not economically viable, such as remote or leisure-dominated markets (Dobruszkes 2006, Chung and Whang 2011, Antunes et al. 2020). Other contributions have highlighted

the important role of other airline products, such as higher frequencies and better timetables, as a key driver to both capturing demand over competition and stimulate new traffic.

In recent years, a better understanding of airline planning problems, along with advancements in data availability and forecasting methodologies, has motivated scholars and practitioners to put greater emphasis on endogenizing demand into network planning models in order to achieve and exploit a more realistic representation of the so-called supply-demand interactions (e.g., Hsu and Wen 2003, Lohatepanont and Barnhart 2004, Dong et al. 2016, Cadarso et al. 2017).

Properly incorporating supply-demand interactions leads to additional challenges. First, it requires the empirical estimation of demand models, which comes with data availability and model estimation issues. Furthermore, the incorporation of supply-demand interactions into mathematical models requires either iterating among sub-models (e.g., Hsu and Wen 2003) or addressing nonlinearities that arise from nonlinear functional forms of advanced demand models (e.g., Cadarso et al. 2017). In both cases, the resulting problem is very hard to solve and requires the development of ad-hoc solution methods. Despite some recent contributions, research on this topic is all but exhaustive and many challenges are still to be addressed.

The aim of this thesis is to contribute to this important stream of research by developing integrated models that explicitly endogenize supply-demand interactions at the various phases of the airline planning process. The thesis is composed of three main chapters, each of which constitutes a stand-alone academic paper and addresses a specific issue related to the modeling of supply-demand interactions in the optimization of air transport networks.

In the next sections, we provide background material on the airline planning process and air travel demand. By these sections, we aim to set a general terminology and familiarize the reader with key concepts that will be recurrently used throughout the thesis. In Section 1.3, we state the thesis main contributions and provide an outline of the remaining chapters.

1.1   Airline Planning Process

The airline planning process is very complex process that comprises many interrelated decisions, spanning from short-term activities of operational nature to medium-term tactical decisions, up to long-term strategic decisions. In the following, we provide a brief overview of the different planning steps, highlighting their main objectives and features, and providing references to relevant developments in the optimization literature.
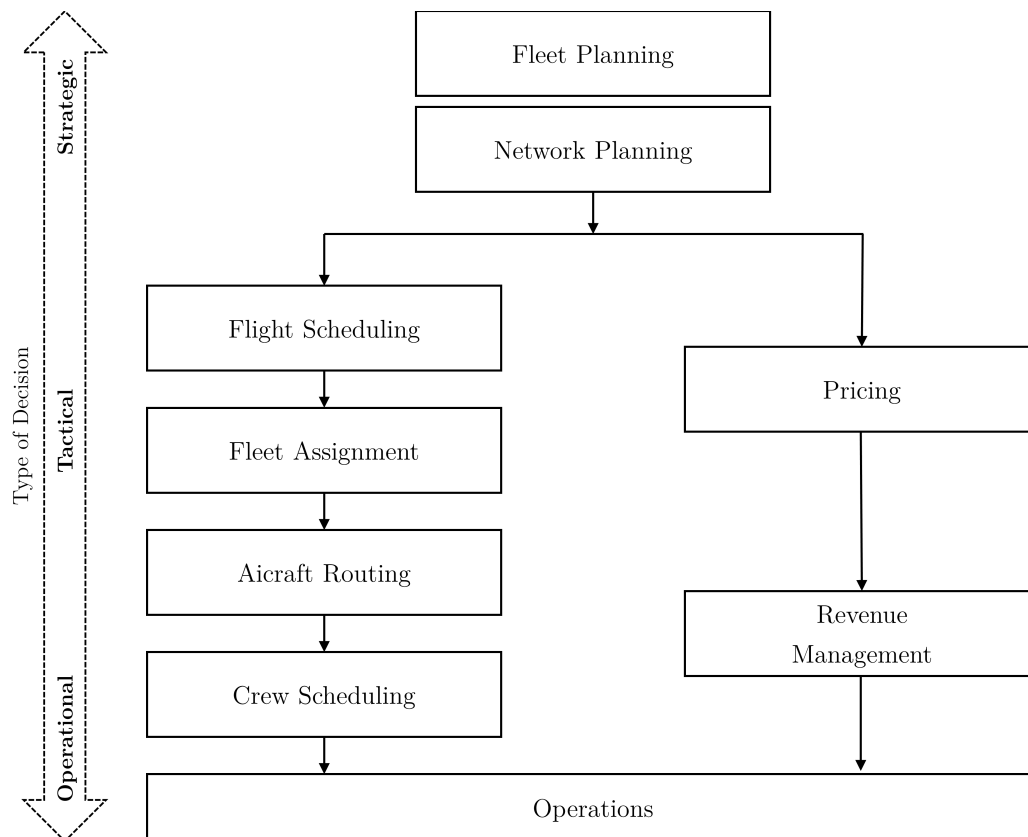
Figure 1-1: Airline planning process (adjusted from Belobaba et al. (2015)

Although represented sequentially (see Figure 1-1), these steps are tightly interrelated, with overlaps and necessary feedback from one stage to another. Traditionally, the sequential representation reflected practice quite accurately. The different planning steps were carried out in sequence and with limited feedback due to organizational constraints and computational reasons. More recently, along with the

advancements in both computing technologies and solution methods, increasing efforts have been put into the joint optimization of different planning stages so as to explicitly capture the interrelationships among them and thus achieve superior decisions.

On the one hand, aiming at integrating all steps of the airline planning process is not (yet) computationally feasible, nor necessarily desirable. Single steps can be very hard combinatorial problems on their own (e.g. aircraft scheduling) and, more importantly, different stages are characterized by different time-horizons and are subject to different levels of uncertainty and data availability. On the other hand, a partial integration of two or three steps has demonstrated significant benefits and has become standard practice—especially at the tactical and operational level.

The first steps that are carried out in the airline planning process are of strategic nature and relate to fleet planning and network design. These decisions are typically made months, or years, before operations and constitute the key inputs for subsequent tactical phases.

Fleet planning

The fleet planning problem consists in defining the fleet size and composition with a long-term outlook. Strategic fleet planning is a crucial task for an airline. Decisions to acquire/lease new aircraft, or dismiss/retire existing aircraft have substantial and long-term impacts on an airline's cost structure, both in terms of investment costs and operating costs. Furthermore, the fleet configuration inevitably affects the ability to serve specific routes profitably and significantly hinder the flexibility of subsequent phases. Fleet decisions are mostly driven by the airline business strategy and "vision", aircraft technical characteristics (e.g., aircraft range, speed, pay-load curve), and financial considerations (e.g., purchasing vs leasing, ownership vs operating costs). Further aircraft selection criteria are represented by (growing) environmental concerns, marketing, and political considerations (Belobaba et al. 2015).

Overall, this leads to a very complex, multi-step, and uncertain decision-making setting, which complicates the development of mathematical models in support of air-

line strategic fleet planning. For this reason, despite the importance of fleet planning and its huge economic impacts, airlines have primarily relied on spreadsheet-based "top-down" approaches when making fleet planning decisions, leveraging aggregate forecasts and KPIs at a macro-regional or sub-network basis, and mostly focusing on the assessment of the more quantifiable economic and financial impacts.

Much research has been done in the field of vehicle fleet planning. A review can be found in Andersson et al. (2010). However, only a few recent works have dealt specifically with the airline industry. These recent contributions in the optimization literature are the works of Hsu et al. (2011), Carreira et al. (2017), Repko and Santos (2017), and Sa et al. (2019), which have proposed more comprehensive optimization models that consider route networks, stochastic demand per origin-destination pair, and detailed cost estimates on a per-flight basis. Collectively, these works constitute a relevant addition to the established toolkit of airline planning currently available and lay the foundations for future research toward the development of data-driven decision support systems.

Network design

The second pillar of airline strategic planning is airline network design, which involves the definition of the network structure, or shape (e.g., hub-and-spoke, point-to-point, or mixed-structures), including the number of hubs and their location, and the identification of the routes to be flown and the markets to be served.

The definition of network structure is intrinsically linked to the airline business model and is a fundamental long-term decision. In the literature, the hub location problem has been extensive research. Starting from the seminal paper from O'Kelly (1987), who presented the first mathematical formulation of the problem applied to passenger airlines, further works have focused on developing solution methods and addressing many problem variants, making transport-oriented hub location a thriving area for operations research (see e.g., Bryan and O'Kelly 1999, Marianov and Serra 2003, Adler and Hashai 2005, Campbell and O'Kelly 2012, Adler et al. 2018, Soylu and Katip 2019). Despite the emphasis in the literature, in practice hub location is not

a very recurrent issue. Given the huge costs and strategic considerations associated with the establishment of hub operations, airlines are very unlikely to ever change the location of their main hub(s). On the contrary, airlines regularly need to revise their flight networks in response to demand fluctuations and to promptly exploit business opportunities due to changes in the market environments. In deciding the best routes to be flown, route evaluation is dominated by economic considerations, which require an in-depth assessment of route profitability based on accurate inputs, including demand and revenue forecasts for the period under consideration.

In practice, route profitability assessment has been mostly supported by "bottom-up" approaches that evaluate each route's profitability individually and incrementally. Models of this kind consider very detailed inputs, including traffic forecast and cost estimates that also leverage information from more tactical decisions, like schedules and flight frequencies. These models thus allow airlines to incrementally evaluate new routes, given a set of route candidates and estimated demand, subject to fleet and capacity constraints. As a main drawback, these approaches fail to suitably capture network interdependencies in hub-and-spoke networks (such as the contribution of connecting passengers) and exploit potential synergies with fleet development strategies.

In an effort to capture these aspects, some contributions in the literature have developed integrated optimization-based approaches for airline network plannning (Teodorović et al. 1994, Jaillet et al. 1996, Hsu and Wen 2000, 2002, 2003, Wen and Hsu 2006) that simultaneously design the network, determine route frequencies, and accommodate fleet sizing considerations. However, despite route profitability plays a central role in the airline business, the literature on integrated models is very scant—especially compared to the attention received by airline hub location problems. More importantly, current models suffer from a number of simplifications and scalability issues that limit their applicability as effective decision support tools in realistic airline networks. This aspect will be further elaborated in Chapter 2.

Given the decisions about the fleet and the routes to be flown, the next steps along the airline planning process involve flight scheduling and fleet assignment. Flight

schedules and aircraft rotation plans are developed up to 1 year in advance and finalized 2–6 months before departure.

Flight scheduling

Flight scheduling, also referred to as schedule design, involves the development a profit-maximizing schedule, defining an origin, a destination, a departure time, and an arrival time for each flight leg. Flight scheduling encompasses two main sub problems: frequency planning and timetable development. Frequency planning determines the optimal service frequency for each route in the airline network subject to aircraft availability, and timetable development determines the specific departure time and arrival time for each flight leg, taking into account flow balance and practical considerations to ensure feasible aircraft routings.

The decisions made at this stage determine airlines' economic performance and bind future decisions to a large extent. Flight frequencies and timetables are acknowledged to be among the most important determinants of passenger demand—along with prices and other factors affecting the quality of service. The determination of flight schedules directly impacts the convenience for passengers, which, in turn, significantly determines the demand and revenues that can be achieved. Hence, the success of flight scheduling optimization is mostly grounded on the availability of demand models that carefully appraise passenger demand for an airline's schedule and the resulting revenues. This calls for accurate disaggregate demand models that account for multifaceted quality of service factors and suitably appraise the offerings by competing airlines. To address this issue, literature has developed a number of forecasting methodologies in support of schedule development practices (Jacobs et al. 2012). Departing from earlier models of airline market share based on an "S-curve" relationship (Button and Drexler 2005, Vaze and Barnhart 2012, Belobaba et al. 2015, Pita et al. 2013), airlines and practitioners have recently leveraged discrete choice methods to estimate itinerary choice models, which replicate how individuals select among air-travel alternatives within a city-pair market (see Section 1.2, for details). Among the relevant itinerary features, previous studies have investigated the role of

various level-of-service, connection quality, and carrier attributes, underscoring the key role played by airfare, travel and connecting time, number of stops, frequency, punctuality, and time-of-day preference (see e.g., Coldren et al. 2003, Coldren and Koppelman 2005, Koppelman et al. 2008, Lurkin et al. 2018). Despite the wide literature on schedule development, the development of suitable demand models in support of flight scheduling—and their integration with the optimization framework—is still a hot topic attracting continuous research attention. Chapters 3 and 4 delve into this aspect more deeply.

Fleet assignment

Once the flight schedule has been defined, the next step in the airline planning process is fleet assignment. Given the number of aircraft of each type, as well as the flight schedule, fleet assignment assigns an aircraft type to each scheduled flight to optimally match passenger demand. This involves trading off passenger revenues and aircraft assignment costs to maximize operating profits, subject to aircraft availability and flow balance constraints across the entire network. This assignment, obviously, becomes more crucial with the increase in the number of different aircraft types, as commonly the case with modern aircraft fleets.

Fleet assignment has been widely researched in the literature and is one of the first tasks that has been routinely supported by optimization tools since their early developments. The first mathematical model of fleet assignment is traditionally traced back to Abara (1989a), who provided a connection-based network formulation able to deal with realistically sized fleet assignment problems. Next, Hane et al. (1995) formulated the problem as a multicommodity time-space network flow model, which has been established as the method of choice in formulating the fleet assignment problem. Building on these contributions, several later works have advanced the leg-based fleet assignment models (FAM) in several ways. Enhanced models have achieved a greater realism and wider acceptance among airlines by considering additional features of real-world airline networks such as itinerary-based demand (IFAM) (Barnhart et al. 2002), flexible departure time (Levin 1971, Rexing et al. 2000), and re-fleeting mecha-

nisms (Sherali et al. 2005). A comprehensive review of airline fleet assignment models can be found in Sherali et al. (2006).

Despite the typical sequence that characterizes flight scheduling and fleet assignment in the planning process, these two decisions are clearly strongly intertwined. On the one hand, flight schedules represent the key input for the fleet assignment, but, on the other hand, aircraft technical specifications, seat capacity, and availability directly affect the optimality of flight schedules. Hence, solving these problems sequentially and separately may result in suboptimal decisions, or even yield infeasibilities.

This consideration has motivated a number of contributions to move from standalone flight scheduling and fleet assignment models toward the simultaneous integration of these two planning stages into the so-called integrated flight scheduling and fleet assignment models (Lohatepanont and Barnhart 2004, Sherali et al. 2010, 2013, Pita et al. 2013, Atasoy et al. 2014, Pita et al. 2014, Dong et al. 2016), with significant benefits.

Aircraft routing & Crew Scheduling

After the fleet assignment is solved, aircraft routing and crew scheduling problems are addressed.

The aircraft routing, or tail assignment problem, determines which flight legs should be optimally allocated to individual aircraft. The sequence of flight legs assigned to each aircraft constitutes an aircraft routing, or rotation, and must satisfy a number of operating constraints. Among others, flights in an aircraft routing must be sequential in time and space, as well as ensure sufficient turnaround times and adherence to maintenance restrictions. The objective function aims at minimizing the cost of allocating aircraft to routings.

Closely related to aircraft routing is the crew scheduling problem. Crew scheduling is about to determine the assignments of crews, namely pilots and cabin crew (flight attendants), to each flight leg in the airline's schedule. A crew schedule specifies the sequence of flight legs and accounts for other activities that each crew member should execute over a period of time, typically spanning one month. The crew scheduling

problem is divided into two sub-problems. First, the crew Pairing problem allocates crews to sequences of flights to minimize crew costs—e.g., pay & credit compensation, overnight stays— while ensuring that all flights are covered by a crew. The outcome is a feasible allocation of anonymous crews to flight strings (pairings) that start and end at the crew's base and satisfy regulatory requirements, such as the 8-in-24 rule or the maximum time away from the base (TAFB), as well as other constraints given by collective agreements between the airline and its employees. Second, the crew rostering problem allocates the resulting pairings to individual crew members, taking into account their qualifications, rest periods, and their specific requests.

The airline aircraft routing and crew planning problems have been subject to extensive research. Mathematically, these problems are typically modeled using a set partitioning formulation. Although simple and linear in form, the number of feasible partitions (crew pairing or aircraft rotations) for a real-world airline network is in practice very high, leading to large-scale combinatorial optimization problems (involving thousands of constraints and billions of variables) that are extremely difficult to solve (Barnhart et al. 2003a). Accordingly, a great deal of research around these topics has been dedicated, on the one hand, to embrace the intricacies and specificities of real-world decision making and, on the other hand, to seek advanced solution methods, such as branch and price (Barnhart et al. 1998b), to effectively address large instances characterizing real-world airline networks. Excellent reviews are provided by Barnhart et al. (2003b) and Kasirzadeh et al. (2017) for the crew scheduling and Gronkvist (2005) and Marla et al. (2018) the tail assignment problem, respectively.

Similar to flight scheduling and fleet assignment, also aircraft routing and crew scheduling are intricately interwoven with themselves and prior phases. For instance, solving sequentially the fleet assignment problem and aircraft routing problems can lead to maintenance-infeasible rotations. The same issue can occur between aircraft routing and crew scheduling. Furthermore, besides the risk of generating infeasible solutions, solving aircraft rotation first and crew scheduling later, or vice versa, may limit the set of feasible crew pairings (or, respectively, the number of feasible rotations), thus reducing optimization flexibility and resulting in potentially suboptimal

solutions.

To overcome these issues, researchers have proposed integrated models that simultaneously optimize fleet assignment and aircraft routing (e.g., Haouari et al. 2011, Liang and Chaovalitwongse 2013), fleet assignment and crew scheduling (e.g., Barnhart et al. 1998c, Gao et al. 2009), aircraft routing and crew scheduling (e.g., Cordeau et al. 2001, Barnhart and Cohn 2004, Mercier et al. 2005, Mercier and Soumis 2007) or the three stages simultaneously (e.g., Sandhu and Klabjan 2007, Papadakos 2009). Due to the hardness of the resulting problems, most of the contributions have relied on heuristic approaches to solve these problems in real-world applications. More recently, Cacchiani and Salazar-González (2017) have proposed exact algorithms to solve the three stages simultaneously and demonstrated their applicability for a regional carrier.

Pricing and revenue management

In parallel to schedule planning, airlines must also address pricing and revenue management decisions. In the air transport industry, pricing refers to the determination of the different fare levels to make available to customers in each market, while revenue management consists in optimizing the number of seats on each flight to be made available at each fare level, limiting low-fare seats and protecting seats for later-booking, higher-yield passengers (Barnhart et al. 2003a). Pricing decisions depend on many considerations, including cost, demand, and service-related decisions that directly depend on the strategic vision and positioning pursued by the airline company. A comprehensive research overview on this topic topics is provided by McGill and Van Ryzin (1999) and Strauss et al. (2018).

Pricing information is requested at all stages of the airline planning process, although at different levels of aggregation. Prices directly affect the demand estimates and are necessary to pursue profit maximization objectives at the different strategic and tactical decision-making stages. Hence, the yield management process is continued throughout the planning process, until the day of operation, to make sure that the seat allocation matches future demand most profitably. Although some contribu-

tions have attempted to integrate pricing and scheduling decisions (see e.g., Atasoy et al. 2014, Barnhart et al. 2009, Dong et al. 2016, Yan et al. 2020)—which surely represents an interesting feature for future research—this is not the current state of practice and is out of the scope of this thesis.

1.2 Demand modeling in air transportation

Air travel demand is a key component of air transport planning. Literature on air travel demand estimation is widespread and concentrates around to two main issues: the identification of main air travel determinants and the development of model formulations that best reflect passenger behaviors, consistently with the level of data aggregation and information available at the different planning stages. Jorge-Calderón (1997) provides a well-accepted classification of air travel determinants into two broad categories: geo-economic characteristics of the territories in which travel occurs and service-related variables, which depend on the air services and are thus partially under control of airlines and airports. Evidence in the literature (e.g., Jorge-Calderón 1997, Grosche et al. 2007, Valdes 2015, Abed et al. 2001, Jankiewicz and Huderek-Glapska 2016, Boonekamp et al. 2018, Adler et al. 2018) have highlighted the role of population, income (GDP per capita), trade and tourism flows, as well as cultural, ethnic and political links between countries, and aviation-dependent employment as key geo-economic determinants. On the other hand, regarding service-related variables, flight frequency, travel time, itinerary routing — proxied by a number of variables capturing the disutility of connecting vs nonstop itineraries, such as connecting time and number of stops—, schedule delay and departure time, price, and punctuality (or stochastic delay) have been identified as key level-of-service attributes driving passenger choices (e.g., Abrahams 1983, Coldren et al. 2003, Coldren and Koppelman 2005, Koppelman et al. 2008, Garrow et al. 2016, Lurkin et al. 2017, 2018).

An important distinction to consider when dealing with the estimation of air travel demand is between demand generation and demand allocation models (Figure 1-2). Demand generation generally refers to the estimation of total travel demand

between two areas (e.g. o and d in Figure 1-2), while demand allocation concerns the redistribution of this demand over the available travel alternatives (e.g. k_1, k_2, and k_3 in in Figure 1-2). Demand generation models are also referred to as O-D market demand models, while demand allocation models as itinerary market share models (Belobaba et al. 2015).

Demand generation **Demand allocation**
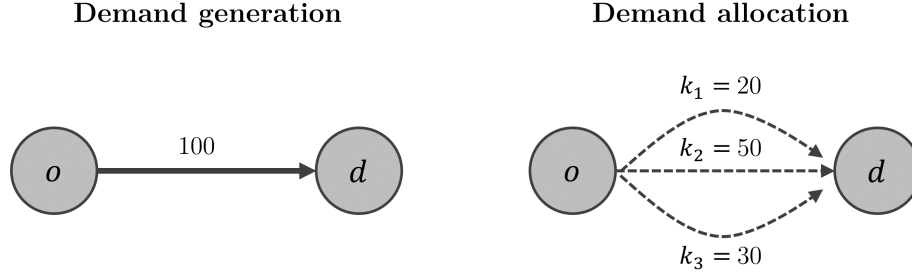
Figure 1-2: Demand generation and demand allocation

Demand generation models usually rely on a gravity formulation (e.g. Jorge-Calderón 1997, Grosche et al. 2007, Adler and Hashai 2005, Adler et al. 2018). Based on an analogy with Newton's law of gravity, these models represent demand as a result of conflicting forces of attraction and impedance. In its simplest form, the gravity model can be stated as follows:

$$D_{od} = \frac{M_o^{\beta_1} M_d^{\beta_2}}{T_{od}^{\beta_3}} \tag{1.1}$$

where D_{od} indicates the total demand between o and d in a given period, M_o and M_d represent attraction factors (e.g. population, gdp) at the origin and destination, and T_{od} includes impedance factors that forbid, or restrain air travel between the two areas (e.g., distance). More advanced formulations have extended the basic gravity model by including service-related variables to appraise the role of service quality in stimulating air travel. Previous studies have considered single factors separately, such as average frequency, average price, or the presence of low-cost carriers. An alternative approach consist in computing the *generalized cost of travel* to combine in a unique variable the multifaceted aspects and components that characterize a typical trip by an air passenger—e.g, travel time, schedule delay, waiting time (Abra-

hams 1983, Lieshout et al. 2016). Demand generation models have been extensively used by practitioners to estimate aggregate traffic flows between territories. However, although disaggregate models can be estimated by segmenting the dataset, for example, by passenger types (business vs leisure), lengths of haul, or airlines, the use of an aggregate multiplicative specification is not very appropriate to capture competition among air travel itineraries.

As anticipated in Section 1.1, airlines have intensively relied on demand allocation models to support tactical decisions regarding service design and capacity allocation. By taking total demand as given, these models estimate the proportion of total market demand that is captured by the different alternatives available in the market. This is achieved by identifying key factors driving the allocation of passengers, collecting this information for all alternatives available, and applying a function that redistribute passengers proportionally to the relevant factors identified. Each step n its own is not trivial and poses substantial theoretical and empirical challenges.

The earliest models of demand allocation in the air travel industry relied on a "S-curve" relating flight frequencies with itinerary market shares (e.g., Button and Drexler 2005, Pita et al. 2013). Let k be a given itinerary operating in market (o, d), demand D_{odk} on itinerary k is estimated as follows:

$$D_{odk} = D_{od} \frac{N_k^{\mu}}{\sum_{k' \in K} N_{k'}^{\mu}} \tag{1.2}$$

where D_{od} is the total market demand between o and d, N_k is the service frequency of itinerary k, K is the set of itineraries available in market (o, d), and μ is an empirical coefficient.

More recently, researchers have been relying on discrete choice modeling to account for more quality factors (rather than only frequency) and improve the degree of behavioral realism in the allocation of passengers over available itineraries. The theory of discrete choice modeling is rooted in the random utility theory, according to which passengers select their preferred itinerary in a given market based on the relative utilities provided by all available itineraries. Utility is intended as an abstract

measure that quantifies the satisfaction received by an individual from consuming a good or service. Since it is unrealistic assuming that the modeler can collect and account for all factors affecting passengers' choice behavior, the utility is modelled as the sum of a deterministic, or observable, component (which can be estimated by the modeler and is typically represented as a linear function of quantifiable relevant choice attributes) and a random component, which is unobserved and unmeasurable by the modeler. According to the assumption on the distribution of the error term, different model specifications are obtained (further details can be found in Ben-Akiva and Lerman (1985) and Train (2003)). In particular, the assumption that the random terms are independently and identically distributed following an extreme-value (or Gumbel) distribution leads to the formulation of the basic multinomial logit (MNL) (McFadden 1974)—which has been widely used in modeling air passenger choices (Coldren et al. 2003, Garrow 2016, Cho et al. 2015). Using the same notation as in Equation 1.2, the MNL formulation is as follows:

$$D_{odk} = D_{od}\frac{e^{V_k}}{\sum_{k' \in K} e^{V_{k'}}} \tag{1.3}$$

where V_k is the deterministic utility component of itinerary k, usually represented as a linear combination of relevant choice factors, i.e., $V_k = \boldsymbol{\beta}^T \boldsymbol{x_k}$ (where $\boldsymbol{x_k}$ is the vector of explanatory variables for itinerary k and $\boldsymbol{\beta}$ is the vector of coefficients to be estimated).

The MNL has recently become the standard for capturing supply-demand interactions in airline tactical planning (see, e.g. Talluri and van Ryzin 2004, Atasoy et al. 2014, Dong et al. 2016). The great popularity of MNL models stems from its flexibility to simultaneously and consistently account for many factors (included in the deterministic component). Another important feature is that MNL models produce a (nonlinear) closed-form expression for choice probabilities, which can be explicitly embedded into optimization models (although not trivially). However, MNL suffers from two main limitations. First, the assumption that error terms are independently distributed forbids MNL models to properly accommodate correlations among alter-

natives. This issue is formally known as *Independence from Irrelevant Alternatives (IIA)* and, when violated, may lead to unrealistic substitution patterns (see Ben-Akiva and Lerman (1985) and Mokhtarian (2016), for details). Second, MNL models estimate unique and deterministic coefficients, thus overlooking any form of taste heterogeneity across individuals. Recent contributions in the field of discrete choice modeling have proposed advanced model formulations to overcome these important limitations of MNL models. *Nested logit models (NL)* extend the basic MNL formulation by grouping alternatives into nests, thus explicitly capturing some (known) forms of correlation among alternatives (Carrasco and de Dios Ortúzar 2002, Coldren and Koppelman 2005, Garrow et al. 2016). *Latent class models (LC)* explicitly address the heterogeneity issues by estimating separate sets of coefficients for homogeneneous groups of respondents (Greene and Hensher 2003, Wen and Lai 2010). *Mixed logit models (MMNL)* adress both issues by the estimation random coefficients (Hensher and Greene 2003, Train 2003, Birolini et al. 2019). Despite the recognized superiority of these models to represent passenger behaviour, their application and systematic incorporation into airline optimization-based decision support tools is very limited.

Demand generation and demand allocation have been traditionally solved sequentially—assessing demand generation at one level of aggregation and then distributing the estimated volumes to lower-level components (Hsiao 2008). By assuming that the total demand is fixed for the allocation model, the sequential approach fails to fully appraise the specificities of single itineraries and how changes in their attributes can impact the overall air traffic volume. Acknowledging this limitation, Wei and Hansen (2005) and Hsiao and Hansen (2011) have proposed similar approaches to tackling the two stages simultaneously based on hierarchical demand models. The proposed hierarchical structure involves (at least) two levels. At the bottom level, an allocation model is used to allocate demand based on the utility values of the different air travel alternatives. At the upper level, the overall utility to travel by air in a given market (compared to the alternatives of not traveling at all or traveling by other modes) is obtained as the sum of exogenous factors (e.g. geo-economic factors) and the composite utility provided by the inner air travel alternatives. This mod-

eling practice achieves the goal to make total market demand elastic to the quality of air transport supply and, more importantly, to estimate the elastic response as a function of the same service-related attributes considered in the demand allocation model. Similar to the capturing of interrelationships among airline planning steps, the integrated modeling of demand generation and allocation constitutes an important aspect of supply-demand interactions that has been substantially underrepresented in the literature.

1.3 Research outline and contributions

This thesis follows the "three paper" format. The first contribution (Chapter 2) deals with airline network planning. The treatment of supply-demand interactions has been largely overlooked at this stage of airline planning. We fill these gap by developing a data-driven optimization model that simultaneously optimize key strategic decisions regarding flight network and fleet composition and explicitly incorporate demand-supply interactions related to flight frequencies and hub-and-spoke operations. The second contribution (Chapter 3) deals with airline tactical planning. The key issue here is to incorporate both demand allocation and generation dynamics into the mathematical modes used to optimize flight scheduling and fleet assignment. To this aim, we rely on a hierarchical demand model to jointly endogenize demand generation and allocation, and develop a solution approach based on piecewise linearization to address the resulting nonlinearities in the mathematical model. The third contribution (Chapter 4) addresses a fundamental issue of air travel demand estimation. This relates to the capturing of correlations between overlapping air transport markets in the estimation of air travel demand. To address this issue, we propose an integrated origin-based demand model that assumes saturation at the origin level—instead of assuming that demand occurs independently within city-pairs—and demonstrate how this can help practitioners in better assessing the impact of changes in air transport supply. We now introduce each chapter more in detail.

Chapter 2: *Airline Network Planning: Data-driven Optimization with Demand-*

supply Interactions

Airlines routinely use analytics tools to support flight scheduling, fleet assignment, revenue management, crew scheduling, and many other operational decisions. However, decision support systems are less prevalent to support strategic planning. This paper fills that gap with an original data-driven optimization model, named *Airline Network Planning with Supply and Demand interactions (ANPSD)*. The ANPSD optimizes network planning (including route planning, flight frequencies and fleet composition), while capturing interdependencies between airline supply and passenger demand. We first estimate a demand function as a function of flight frequencies and network composition, using a two-stage least-squares procedure fitted to historical data. We then formalize the ANPSD by integrating the empirical demand function into an optimization model. The model is formulated as a non-convex mixed-integer program. To solve it, we develop an exact cutting plane algorithm, named 2αECP, which iteratively generates hyperplanes to develop an outer approximation of the non-linear demand function. Computational results show that the 2αECP algorithm outperforms state-of-the-art benchmarks and generates tight solution quality guarantees. Case study results based on the network of a major European carrier show that the ANPSD provides much stronger solutions than baselines that ignore—fully or partially—demand-supply interactions.

This paper was co-authored by Alexandre Jacquillat (Sloan School of Management at the Massachusetts Institute of Technology,), António Pais Antunes (University of Coimbra), and Mattia Cattaneo (University of Bergamo), and is currently under submission in *Transportation Research Part B: Methodological.*

Chapter 3: *Integrated flight scheduling and fleet assignment with improved supply-demand interaction modeling*

Flight scheduling and fleet assignment are important steps of an airline planning process. In light of the reciprocal relationship between air transport supply and demand, a key element of these models is to devise effective methods to both incorporating estimation of total market demand and allocating passengers over the

available itineraries in a specific market. In this paper, we present a novel mixed-integer nonlinear flight scheduling and fleet assignment optimization model wherein air travel demand generation and allocation are simultaneously and consistently endogenized. Using a nested logit formulation, we jointly model competition among air travel itineraries and appraise the contribution of specific itinerary attributes to demand generation, therefore yielding a more comprehensive and explicit representation of supply-demand interactions. Computational testing based on realistic problem instances reveals that the model can optimize mid-size hub-and-spoke networks within reasonable time. Further analyses illustrate the benefits that can be derived from the application of the proposed approach using real-world data for a major European airline. Results demonstrate that the proposed approach can significantly enhance operating profits by up to 6.9% and better reveal opportunities for demand stimulation against a conventional approach using inelastic trip generation.

This paper was co-authored by Prof. António Pais Antunes (University of Coimbra), Mattia Cattaneo (University of Bergamo), Paolo Malighetti (University of Bergamo), and Stefano Paleari (University of Bergamo), and is currently under submission (R&R) in *Transportation Research Part B: Methodological*.

Chapter 4: *Integrated origin-based demand modeling for air transportation*

This paper proposes an origin-based approach to the estimation of the demand for air travel. Whereas the prevailing approach in the literature involves the independent estimation of air flows between city-pair markets, the proposed framework explicitly accounts for the determinants of outbound trips and substitution patterns between destinations. We simultaneously integrate demand generation and allocation using a multilevel aggregate nested logit formulation that covers the choices of whether or not to travel by air, where to travel (destination), and how to travel (itinerary). Two specifications are proposed to reflect systematic differences between lengths of haul and the bootstrap is applied to jointly address endogeneity issues and data missingness. The validity of the proposed approach is tested over the entire network of outbound air trips from Italy in 2018. Results highlight systematic differences

in sensitivity to travel determinants between medium- and long-haul travel itinerary choices. Additionally, nesting coefficients demonstrate the validity of the origin-based approach, as attribute changes in air transport supply are found to significantly affect both passenger distribution among destinations and trip making.

This paper was co-authored by Mattia Cattaneo, Paolo Malighetti, and Chiara Morlotti, all affiliated at the University of Bergamo. It was awarded the best PhD student paper award at the 2019 ATRS World Conference in Amsterdam and was recently published in *Transportation Research Part E: Logistics and Transportation Review.*

Finally, in Chapter 5, we conclude the thesis and discuss directions for future research.

Chapter 2

Airline Network Planning: Data-driven Optimization with Demand-supply Interactions

2.1 Introduction

The airline industry has been facing rapid growth globally for the past decades. Yet, airlines have had to react to a number of shocks, such as the crisis following the terrorist attacks of September 11, 2001 and the financial downturn in 2008-10. Even more dramatically, the Covid-19 pandemic put the entire industry to a brutal halt in 2020, resulting in downsized operations at hub airports, global cuts in service frequency, and early aircraft retirements—raising critical questions for airlines to ramp operations back up as society progressively reopens. In addition to these global patterns, individual airlines are often subject to demand fluctuations on specific routes. These ever-changing patterns in air travel demand create massive challenges and opportunities for airlines to revise their flight networks and fleets in response to long-term outlooks on the origin-destination markets.

The airline planning process comprises many interrelated decisions, spanning strategic decisions (e.g., route planning, fleet planning), tactical decisions (e.g., sched-

ule planning, fleet assignment, revenue management) and operating decisions (e.g., aircraft routing, crew rostering, schedule recovery). Collectively, these decisions have a tremendous impact on airline operating profitability. Airlines have long been at the forefront of innovation in operations research and, today, tactical and operational decisions are routinely supported by dedicated optimization software. For instance, a new generation of fleet assignment algorithms in the late 1980s and early 1990s resulted in a 1.4% improvement in operating margins at American Airlines (Abara 1989a) and annual savings of $100 million at Delta Airlines (Subramanian et al. 1994). A decade later, schedule planning software again resulted in significant profit improvements, estimated at $500 million a year at American Airlines (Barnhart and Cohn 2004). Airlines have achieved similar success in revenue management, schedule recovery, crew planning, and other tactical and operational areas.

In contrast, decision support systems are less prevalent in airline strategic planning. The academic literature has mainly focused on hub location problems (see, e.g., Jaillet et al. 1996, Marianov and Serra 2003, Soylu and Katip 2019). However, in practice, the location of hub airports has remained all but fixed for the vast majority of carriers since their inception. For example, British Airways' boardrooms are rarely preoccupied by whether they should keep a hub in London or relocate it to Amsterdam. In comparison, airlines regularly need to re-appraise which routes to open and close, which origin-destination markets to serve, and which aircraft to operate. The optimization literature is more sparse in these areas. In practice, many strategic decisions, despite being key drivers of airline operations, are still rarely supported by analytical tools.

To fill this gap, we formulate and solve a data-driven optimization model in support of airline strategic planning, referred to as *Airline Network Planning with Supply and Demand interactions (ANPSD)*. The model considers the airline's hub airports as fixed. It optimizes network planning decisions to maximize operating profits, subject to operating constraints. As such, the model adopts a timeframe spanning a medium-to long-term horizon (months to years). The outputs of the model are the core inputs of the subsequent tactical phase of the airline planning process.

A key challenge lies in capturing the interdependencies between airlines' networks of flights, on the supply side, and passenger flows, on the demand side. On the one hand, the optimal network of flights obviously depends on passenger demand. Vice versa, the quality of an airline's offerings on each origin-destination market impacts passenger demand through demand generation (by stimulating new traffic) and demand capture (by attracting passengers from competitors). In order to capture these interdependencies, we first estimate passenger demand from historical passenger flow data, as a function of demand-side variables and supply-side variables. We then embed this demand model into an optimization model to support network planning decisions, and develop an original mixed integer non-linear programming algorithm to solve it.

Specifically, this paper makes the following contributions:

1. *It develops a data-driven model to estimate passenger demand, as a function of airline network characteristics.* We use a two-stage least-squares procedure to estimate a leg-based gravity model from historical passenger flow data as a function of demographic, geographic and economic variables, as well as supply-related variables. In particular, we let demand vary with flight frequencies, to capture the positive effect of frequency on demand, and with network structure, to capture the connectivity effects originating from the development and consolidation of hub-and-spoke networks. As a result, this empirical model explicitly captures the interdependencies between supply-side network planning and demand-side passenger flows.

2. *It provides a novel optimization model to support airline strategic planning that captures demand-supply interactions.* We develop an original formulation that optimizes network planning, which we define as encompassing (i) route planning, i..e., on which origin-destination pairs to operate, (ii) frequency planning, i.e., how many flights to operate on each route, and (iii) fleet composition, i.e., which aircraft types to operate and where. We capture demand-supply interdependencies by integrating our empirical model of passenger demand into the optimization. From a technical standpoint, however, these demand-supply in-

teractions create constraints of the form $D = \gamma f^u s^v$ (where D is the demand, f is the flight frequency, s is the number of spokes in the network). As a result, the ANPSD is formulated as a non-linear, non-convex mixed-integer program.

3. *It develops an exact cutting-plane algorithm based on outer approximation to solve the resulting non-linear, non-convex mixed-integer optimization model, referred to as 2αECP.* Leveraging the structure of the demand function, we define, in any point, two semi-hyperplanes that yield a valid relaxation of of the true (non-linear, non-convex) function. Accordingly, we design a gradient-based procedure that extends outer approximation schemes from convex to non-convex problems, using two semi-hyperplanes as opposed to a single one. At each iteration, the algorithm solves mixed-integer linear programming models to update a feasible solution (lower bound to a maximization problem) and a solution guarantee (upper bound), until convergence to a provable optimality gap. Ultimately, the algorithm provides a new solution approach to a broad class of data-driven optimization problems that consist of (i) a mixed-integer optimization problem and (ii) a non-convex function stemming, for instance, from *log-log* regression specifications ($\log y = \gamma + \boldsymbol{\beta}^T \log \boldsymbol{x}$).

4. *It shows that the proposed algorithm generates high-quality solutions in short computational times, outperforming state-of-the-art benchmarks.* We consider baseline solution methods based on discretization and linearization, using traditional convex combination techniques (Keha et al. 2004, Lee and Wilson 2001, Padberg 2000) and the more recent logarithmic bivariate piecewise linearization (Vielma et al. 2010, Vielma and Nemhauser 2011). We conduct a comprehensive computational study in a controlled experimental setting, using randomly-generated instances inspired from real-world data. This environment enable us to investigate the scalability of the proposed approach and systematically assess computational performance as a function of network structure—one or two hubs—and network size—100, 250, and 500 routes. Results suggest that our cutting-plane algorithm outperforms the benchmarks, returning better solutions and more consistent performance. Ultimately, the proposed algorithm

generates optimal, or near-optimal solutions, to realistic problem instances in short computational times (2 hours), consistent with practical implementation requirements.

5. *It demonstrates the practical impact of the proposed modeling and algorithmic approach using real-world data, providing insights into the development and consolidation of airline networks.* We apply the model to the continental network of Alitalia, consisting in 2018 of two hubs, 172 monthly routes and 3,028 weekly flights on average. The demand function is fitted to historical data on all intra-European flights in 2018. Results demonstrate that capturing the demand-supply interactions in our data-driven optimization model results in substantial benefits, as compared to baselines that ignore—fully or partially—these interdependencies. From a managerial standpoint, the proposed model provides strategic insights into how to balance competing network planning objectives, such as consolidating operations at major hubs vs. leveraging a two-hub network, and increasing flight frequency and service offers to the busiest airports vs. covering a broader set of destinations. Ultimately, this approach can drive the synergistic evolution of flight networks and aircraft fleets in airline networks.

The remainder of the paper is organized as follows. Section 2.2 reviews the related literature on airline strategic planning. Section 2.3 describes the overall modeling structure, which is further detailed in Section 2.4 (demand model) and Section 2.5 (optimization model). In Section 2.6, we present our cutting plane algorithm and the linearization benchmarks. We report the computational results using synthetic data in Section 2.7 and real-world case study data in Section 2.8. We summarize our contributions and outline future research directions in Section 2.9.

2.2 Literature Review

Within airline strategic planning, the scientific literature has primarily concentrated on hub location and fleet planning problems. The hub location problem involves determining the location of hub airports to connect passenger flows from origin to

destination nodes. O'Kelly (1987) provided the first mathematical formulation of the problem. Follow-up research was devoted to improve the model formulation, develop solution methods, and address different problem variants (see Bryan and O'Kelly 1999, Campbell and O'Kelly 2012, for reviews). The hub location problem is primarily focused on the overall network shape—for instance, hub-and-spoke vs. point-to-point, how many hubs and where. As such, these models do not capture other important decisions in network planning, such as route planning, flight frequency and aircraft operations. The fleet planning problem consists of defining the fleet size and composition with a long-term outlook. In practice airlines have primarily relied on spreadsheet-based "top-down" approaches, leveraging aggregate forecasts and key performance indicators at a macro-regional or sub-network basis, and focusing almost exclusively on financial aspects (e.g., buying vs. leasing aircraft) (Belobaba et al. 2015). Recently, optimization models have been proposed to support when to dismiss/sell vs. acquire/lease aircraft, providing a more systematic and integrated treatment of demand uncertainty, purchasing and leasing alternatives, and operating costs (Listes and Dekker 2005, Hsu et al. 2011, Dožić and Kalić 2015, Carreira et al. 2017, Repko and Santos 2017, Sa et al. 2019). These models do not fully capture route planning considerations, instead considering a restricted set of routes as inputs. Yet, fleet planning and route planning decisions are strongly interdependent, in that route profitability depends on available aircraft and, conversely, fleeting decisions depend on how aircraft will be used in the network of flights and how they will contribute to operating profitability. These interdependencies motivate new research to investigate the route planning and fleet composition problems in an integrated manner.

Another extensive branch of the literature has focused on developing optimization models in support of tactical planning. Notable examples include flight scheduling (Lohatepanont and Barnhart 2004), fleet assignment (Abara 1989b, Hane et al. 1995, Sherali et al. 2006), maintenance aircraft routing (Desaulniers et al. 1997, Barnhart et al. 1998a), and crew scheduling (Barnhart et al. 2003b).

In contrast, this paper focuses on the *network planning problem*, encompassing route selection, frequency planning and fleet composition decisions. As such, the net-

work planning problem stands somewhat in-between the very long-term problems of hub location and fleet planning and the more tactical problems of flight scheduling and fleet assignment. Despite involving critical and recurrent strategic decisions for the airlines, network planning has received limited attention in the literature. Teodorović et al. (1994) first proposed an optimization model to design an airline's network and determine flight frequencies. To cope with the combinatorial complexity of the problem, the authors proposed a two-step approach, which first determines a set of (non-stop or connecting) route candidates, and then determines flight frequencies for this set of routes. Jaillet et al. (1996) introduced three integer programming models to design capacitated networks and routing policies without assuming a given network structure a priori. Due to the difficulty to solve even small instances by exact methods, the authors introduced a series of heuristic algorithms. These two seminal papers have provided important advances; still, they rely on restrictive assumptions. First, the models focus on cost minimization, thus providing a partial appraisal of each route's economic viability. Second, fleet considerations are limited to a handful of generic aircraft types, thus failing to capture the complexities of modern fleets. Third, these models start from predetermined origin-destination markets. As such, the focus is more on identifying the adequate network structure (and supporting frequencies) rather than selecting the most profitable markets.

A key ingredient of the network planning problem is the prediction of passenger demand on each route. Extensive empirical research developed data-driven models to estimate passenger demand, as an input into downstream applications (e.g., downstream optimization models). Notably, Adler and Hashai (2005) an Adler et al. (2018) developed gravity models to estimate passenger flows in developing countries, in support of hub location and network development decisions; Wei and Hansen (2005) proposed a hierarchical demand model to assess competition dynamics in non-stop duopoly markets; and Wei and Hansen (2006) developed an aggregate leg-based demand model to predict demand in a hub-and-spoke environment and evaluate network expansion alternatives.

As mentioned in the introduction, a fundamental challenge in this literature in-

volves the interdependencies between airline supply and passenger demand. These demand-supply interactions have become a common feature of tactical airline planning models, which leverage disaggregate demand models to capture quality competition among air travel itineraries. For instance, Atasoy et al. (2014) and Dong et al. (2016) propose an integrated model to optimize flight scheduling, fleet assignment and pricing, where demand is allocated across itineraries using a multinomial logit formulation. Cadarso et al. (2017) capture demand-supply interactions in optimizing frequency planning and flight scheduling, accounting for inter-airline competition inter-modal competition with high speed rail. Wei et al. (2020) optimize fleet assignment and flight timetabling, by capturing the endogeneity of passenger demand by means of a generalized attraction model. In Chapter 3, we use a nested logit model to capture demand generation and demand allocation dynamics in integrated flight scheduling and fleet assignment models.

In contrast, supply-demand interactions have remained largely overlooked in strategic planning. Existing approaches primarily rely on a fixed matrix of origin-destination flows. The focus is thus primarily on demand stochasticity, captured by means of fuzzy theory Teodorović et al. (1994), grey theory (Hsu and Wen 2000, Wen and Hsu 2006), a reliability evaluation procedure (Hsu and Wen 2002), and two-stage stochastic programming (Yang 2010). One exception is the paper from Hsu and Wen (2003), which iterates between frequency planning and passenger choice modeling. Still, the model relies on pre-selected routes and fixed aircraft fleets.

Finally, tackling the network planning problem—especially with a comprehensive empirical model that captures demand-supply interactions—raises additional questions on computational tractability. Due to the complexity of the problem, existing models have been applied to small/medium-scale networks, with 7 nodes (Teodorović et al. 1994), 10 nodes (Hsu and Wen 2002, 2003, Wen and Hsu 2006), 13 nodes (Hsu and Wen 2000), and 39 nodes (Jaillet et al. 1996). While insightful, these settings remain much simpler than those of real-world airline networks, thus limiting the potential applicability of existing models as effective decision support tools.

This discussion highlights three main limitations of the literature on airline strate-

gic planning: (i) no systematic consideration of network planning decisions (e.g., route selection, flight frequencies, fleet composition); (ii) no systematic consideration of demand-supply interactions, and (iii) limited tractability in real-world large-scale networks. This paper fills this gap by proposing an original modeling framework for comprehensive network planning, by embedding a data-driven passenger demand function that incorporates demand-supply interactions, and by developing an exact algorithm to solve the network planning model in large-scale networks of flights.

2.3 Modeling Framework

We propose in this paper an integrated modeling framework for comprehensive airline network design, referred to as *Airline Network Planning with Supply and Demand interactions (ANPSD)*. This modeling framework comprises two main elements, as depicted in Figure 2-1.
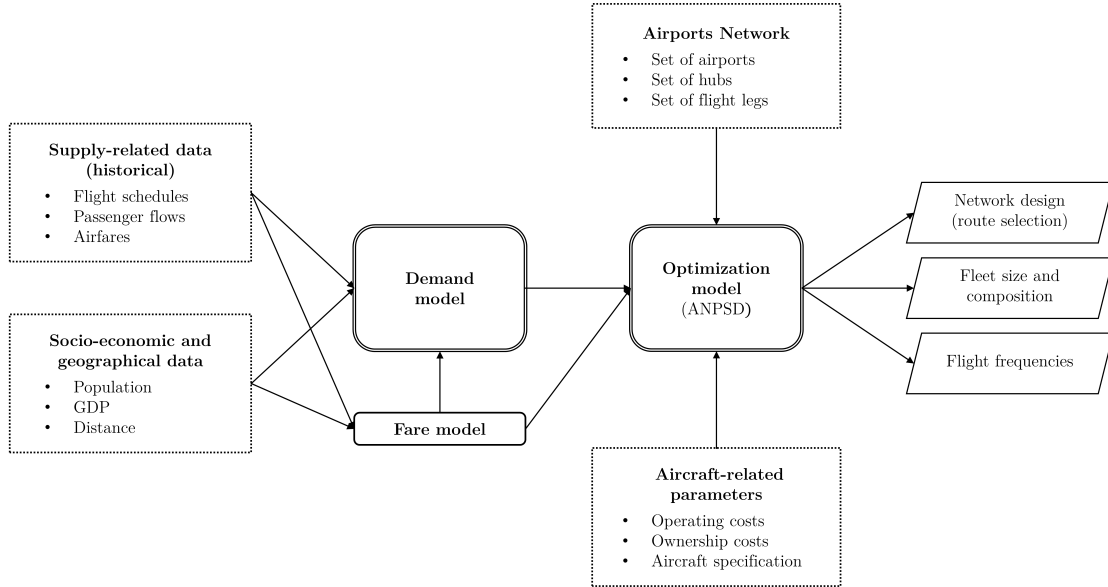
Figure 2-1: Overview of modeling framework.

First, the *demand model* takes as inputs historical data on flight schedules, air-fares, passenger flows, and socio-economic variables. It estimates monthly passenger demand on each flight leg, including both non-stop and connecting passengers, using

a two-stage least squares approach with instrumental variables to capture the endogeneity between demand and flight frequency. Its output is a parametric demand function that captures the dependencies of passenger demand with service frequency and network structure (proxied by the number of spokes).

Second, the *optimization model* takes as inputs the demand model, the set of candidate airports and flight legs in the network, and the set of fleet types. The model optimizes network planning decisions, i.e., route planning, flight frequency, and fleet composition decisions. Due to the form of the data-driven demand function, the ANPSD is formulated as a mixed integer non-linear, non-convex programming model.

In its current form, the ANPSD considers a deterministic demand model to focus on capturing supply-demand interactions in strategic planning optimization. While the proposed modeling framework can be readily used to account for uncertainties through sensitivity analyses, a systematic integration of stochasticity—by quantifying prediction error in the demand estimation model and developing a stochastic optimization formulation accordingly—represents an interesting avenue for future research.

2.4   Demand Model

Let us first note, at the outset, that the proposed integrated approach to airline network planning requires a new empirical framework for passenger demand estimation, as opposed to merely using existing models from the literature. As highlighted in Section 2.2, previous work on airline network planning did not focus on empirical demand estimation with demand-supply interactions. Recent research has embedded disaggregate demand models into tactical optimization models for flight scheduling and fleet assignment. These models typically estimate demand at the itinerary level, often relying on discrete-choice methodologies. The key factor driver passenger demand is itinerary attractiveness, computed as a function of detailed schedule-related attributes, such as total travel time, routing, departure times, and average price.

These attributes are partly or entirely determined by tactical optimization models. In contrast, our network planning problem focuses on route selection and flight frequencies at the more aggregate level, over a longer time horizon. Therefore, itineraries' features cannot be exploited to allocate passengers among competing itineraries, thus requiring a new data-driven approach to estimate passenger flows.

We develop this empirical demand estimation procedure in this section. We first introduce the variables in Section 2.4.1 and the data sources in Section 2.4.2; then, we report and discuss the model estimation (Section 2.4.3) and empirical results (Section 2.4.4).

2.4.1 Model Variables

In defining the variables of the demand model, we index flight legs by i, origin and destination airports by o and d, time periods (months) by t, and airlines by k. Our dependent variable, denoted by D_{it}^k, is the demand on flight leg i for airline k. Demand is defined as the sum of nonstop passengers, incoming passengers (connecting at the origin airport) and outgoing passengers (connecting at the destination airport). We leverage a *gravity formulation*, one of the most successful empirical models to forecast passenger demand (see, e.g., Jorge-Calderón 1997, Adler and Hashai 2005, Grosche et al. 2007, Boonekamp et al. 2018), which includes three sets of predictors: impedance, generative, and service-related variables, including leg-based supply variables and network variables.

Impedance variables characterize the "distance" between the origin and the destination. We define: (i) the great circle distance (thousands of kilometers) between the airports ($dist_i$), (ii) the squared distance ($dist_i^2$), and (iii) the economic distance defined by Adler and Hashai (2005) as the absolute difference between the per-capita income of the catchment areas (Δgdp_i).

Generative variables characterize the passenger pool at each airport based on geo-economic determinants. The market potential is estimated by the product of the population size of the catchment areas at the origin and destination ($pop_o pop_d$) (Grosche et al. 2007). We add the product of the yearly volume of local passengers at

the two airports in the previous year ($ylocal_o ylocal_d$) to account for propensity to fly. We define a variable to capture the degree of seasonality on flight leg i at time t. It is expressed as the sum of standardized seasonality at the origin and destination airports ($local_{ot}^{std} + local_{dt}^{std}$), defined such that each component is positive (resp. negative) if the passenger flow in month t is higher (resp. lower) than the mean yearly value at the airport.[1]

Service-related variables characterize the supply decisions from the airline and the resulting quality of service. We consider the number of flights (f_{it}^k) by airline k on leg i in month t. Similarly, we define a price variable ($price_{it}^k$) as the average of the non-stop fare and the revenue contribution from connecting traffic, assuming proration by distance. A dummy variable is introduced to indicate whether airline k is a low-cost carrier (LCC) (lcc^k=1) or a full-service carrier (FSC) (lcc^k=0). This variable reflects different business model between LCCs and FSCs, and accounts for the demand stimulation effect of LCCs on point-to-point traffic (Bhadra 2003). All these variables capture the non-stop local market between airports o and d. We add the number of spokes (s_{it}^k), defined as the number of airports a such that airline k operates a direct flight from a to the origin o or from the destination d to a. This variable captures the "size" of the network, thus accounting for interdependencies across hub-and-spoke operations (Wei and Hansen 2006). Since LCCs rely much less on hub-and-spoke networks than FSC, network interdependencies are expected to be smaller for LCCs. Accordingly, we interact the number of spokes with the low-cost dummy.

2.4.2 Data Sources

The data for the network variables included in the demand model is taken from the Official Airline Guide (OAG), and concerns Europe in the year 2018. Two data modules are used: (i) the OAG Schedule Analyser, which provides data on flight

---

[1]We use values from the previous year to avoid simultaneity issues in the estimation procedure, and to ensure that we only leverage predictors that are available to decision-makers. This is a minor concern because seasonal patterns do not vary significantly from year to year.

schedules, and (ii) the OAG Traffic Analyser, which provides data on air passenger flows and average fares. To remove thin flight observations, we restrict the dataset to combinations with at least one flight per week and a load factor of at least 30%. The sample comprises 191,160 observations.

The geo-economic variables were computed using a Geographic Information System (GIS). We retrieve airport coordinates from the Open Flights Airport Database. For each airport, we define a circular catchment area with a radius of 100 km, and calculate the population and Gross Domestic Product (GDP) based on high-resolution (30 arcsec) global spatial datasets (see Tatem 2017, Kummu et al. 2018, for details). Table 2.1 summarizes these data.

Table 2.1: Data summary.

Variables	Average	Std dev	Data Range *(5th, 50th, 95th)*
Population ('000)	5,126	4,779	[268, 3,666, 16,412]
Local passengers (yearly) ('000)	12,901	12,000	[377, 8,651, 38,008]
GDP per capita ($)	38,308	13,354	[18,886, 37,387, 60,431]
Distance (km)	1,406	830	[319, 1,278, 3,027]
Price ($)	129	68	[45, 118, 248]
Frequency (montly)	36	49	[4, 18, 124]
Number of spokes	61	48	[10, 48, 152]

2.4.3 Model Estimation

The gravity model we use to describe passenger demand implies a non-linear multiplicative specification of the form $y = \prod_{j=1}^{n} x_j^{\beta_j}$, which is estimated by taking the logs of both sides, i.e., $\log y = \sum_{j=1}^{n} \beta_j \log x_j$, and fitting linear regression models. However, the estimation of this model involves three complexities.

First, we face potential endogeneity between passenger demand and service frequency—higher passenger demand leads airlines to increase service frequency, but service frequency also induces higher demand. In such instances, ordinary least squares (OLS) regression can lead to biased and inconsistent estimates. Accordingly, we adopt a two-stage least squares (2SLS) approach with three exogenous instruments: (i) the average aircraft size used by airline k on leg i ($acsize_{it}^{k}$) (Boonekamp et al. 2018); (ii) the average frequency operated by airline k on flight legs of similar distance

(±300 km) as leg i ($avgfreq_{it}^k$)(Cattaneo et al. 2018); and (iii) the dominance index ($domindex_i^k$), defined as the product of the seat capacity shares of airline k at the origin and destination airports. These three instruments are indeed correlated with flight frequency but have no direct impact on demand.

Second, the fare variable has 31% of missing data. We therefore estimate a supply-exogeneous fare model for imputation, considering three main determinants: (i) operating costs, proxied by the product of distance (thousands of km) and fuel cost (US\$ per gallon) ($distfuel_i$); (ii) the airline business model, captured by the low-cost dummy (lcc^k); and (iii) local competition, proxied by the number of carriers competing with airline k in the local market (N_{it}^k) (Brueckner et al. 1992).[2]

Third, the inclusion of fare variables in the demand estimation is also challenged by the endogeneity between price and passenger demand—higher demand leads airlines to increase prices, but higher airfares also induce lower demand. In our setting, however, we avoid this issue by leveraging the operating cost variable ($distfuel_i$). Indeed, this variable provides a valid instrumental variable, as it is correlated with prices but not with demand (Hsiao and Hansen 2011, Birolini et al. 2020).

Equation (2.1) reports the estimated fare model ($R^2 = 0.27$).

$$p_i^k = 131.34 - 6,59distfuel_i + 3.11distfuel_i^2 - 11.05lcc^k - 8.91N_{it}^k. \tag{2.1}$$

The model's relatively poor fit primarily follows from the noise and high heterogeneity in airfare data—a common issues in air transportation research (e.g., Barnhart and Cohn 2004, Lieshout et al. 2016). Yet, the estimated coefficients (all statistically significant at the 99% confidence level) produce reasonable fare estimates and high-

---

[2]In theory, this imputation procedure could be challenged by the endogeneity between price and competition. However, this endogeneity is not too severe in our case, since the fare variable encompasses several markets, therefore reducing the simultaneity with leg-based demand. More broadly, the endogeneity between price and competition has been long debated in the literature. The proper correction would require the development of a complete structural model (Berry and Reiss 2007), which goes beyond the scope of this paper. An alternative approach is to correct for price-competition endogeneity with instrumental variables. However, the identification of valid instruments is problematic, since they mostly entail route characteristics that might serve themselves as explanatory variables. Ultimately, our approach is consistent with the literature (Brueckner et al. 2013).

light meaningful substitution patterns. In particular, the price elasticity estimate of -0.65 is in line with previous findings (Brons et al. 2002). Finally, it is important to note that, in practice, airlines have access to much better airfare estimates to refine these inputs, based on proprietary data and market research.

2.4.4 Estimation Results

Three different model versions are estimated, referred to as M0, M1 and M2, which capture increasing extents of demand-supply interactions. Specifically, M0 omits the variables characterizing frequency and number of spokes; M1 adds frequency but excludes the number of spokes; and M2 is the full model. Table 2.2 reports each model's estimates. Table 2.3 shows the models' predictive performance obtained by Monte Carlo cross-validation using 100 subsamples.

Note that the coefficients in models M1 and M2 are statistically significant with the expected signs. The first-stage regression coefficients also show the validity of the instrumental variables ($domindex_i^k$, $acsize_{it}^k$, and $avgfreq_{it}^k$), which have the expected signs and are significant.

Specifically, we find that population and local traffic at the market endpoints are positively correlated with passenger demand. Distance exhibits an inverse U-shape relationship, reflecting the effects of intermodal competition on shorter routes and stronger links between less distant areas. The economic distance between origin and destination has a small, but negative effect. Seasonality has a positive impact on demand. We obtain price elasticity estimates of -0.35 (M1) and -0.65 (M2), in line with previous values found in the literature (e.g., Brons et al. 2002). Last, the business model dummy indicates a larger point-to-point traffic for LCCs than FSCs, *ceteris paribus*.

We highlight next two benefits of capturing network variables in M2, as opposed to focusing on service frequency only in M1. First, M2 yields a more reasonable frequency elasticity parameter (0.9018, as opposed to 1.0595 with M1). Indeed, frequency elasticity is expected to be less than one, indicative of diminishing returns. Second, M2 allows to isolate the effect of incoming and outgoing connections (elastic-

ity of 0.1514) and disentangle the effects of hub-and-spoke operations. Most importantly, these results underscore the importance of capturing supply-level variables. Indeed, cross-validation results show that M0 exhibits poor predictive performance, with $R^2_{M0} = 0.37$ and an out-of-sample MAPE of 84%. Adding service frequency greatly increases the model's predictive power—$R^2_{M1} = 0.896$. Moreover, accounting for the number of spokes further improves predictive performance—M2 increases R^2 to 0.906 and reduces out-of-sample MAE by 1.3%—along with consistent prediction power among training and testing samples. Collectively, these results highlight the need of embedding demand-supply interactions into the optimization model.

Table 2.2: Demand parameters (robust standard errors reported in parentheses).

	M0(OLS)	M1(2SLS)	M2(2SLS)	M1(1st Stage)	M2(1st Stage)
$constant$	-0.9980***	3.7606***	5.2397***	-2.3343***	-0.7523***
	(0.1712)	(0.069)	(0.0666)	(0.146)	(0.151)
$\log pop_o pop_d$	0.0932***	0.0226***	0.0212***	0.0440***	0.0392***
	(0.0015)	(0.0007)	(0.0006)	(0.0012)	(0.0012)
$\log ylocal_o ylocal_d$	0.2458***	0.0333***	0.0382***	0.1780***	0.1583***
	(0.0013)	(0.0008)	(0.0008)	(0.0012)	(0.0014)
$dist_i$	-0.8578***	0.5765***	0.3464***	-0.5571***	-0.6075***
	(0.0084)	(0.0053)	(0.0053)	(0.0083)	(0.0086)
$dist_i^2$	0.1440***	-0.0819***	-0.0226***	0.1482***	0.1693***
	(0.003)	(0.0015)	(0.0014)	(0.0026)	(0.0027)
$\log \Delta gdp_i$	-0.0038**	0.0013	-0.0024***	-0.0103***	-0.0106***
	(0.0019)	(0.0008)	(0.0007)	(0.0015)	(0.0015)
$local_{ot}^{std} + local_{dt}^{std}$	0.0233***	0.0116***	0.0078***	0.0276***	0.0216***
	(0.0011)	(0.0004)	(0.0004)	(0.0009)	(0.0009)
$\log p_{it}^k$	-0.1747***	-0.3557***	-0.6492***	-0.6083***	-0.7971***
	(0.0317)	(0.0124)	(0.012)	(0.0276)	(0.0278)
lcc^k	-0.0178***	0.4094***	0.3791***	-0.2263***	-0.4477***
	(0.0047)	(0.0023)	(0.0076)	(0.0042)	(0.0163)
$\log f_{it}^k$		1.0595***	0.9018***		
		(0.0026)	(0.003)		
$\log s_{it}^k$			0.1514***		0.0787***
			(0.0018)		(0.0034)
$\log s_{it}^k lcc^k$			-0.0179***		0.0517***
			(0.0018)		(0.0042)
$domindex_i^k$				1.4341***	1.2006***
				(0.0163)	(0.0173)
$acsize_{it}^k$				-0.0005***	-0.0008***
				(0.0001)	(0.0001)
$avgfreq_{it}^k$				0.6155***	0.5807***
				(0.0034)	(0.0039)

Table 2.3: Monte Carlo sampling validation.

Model	R^2	Mean Absolute Error (MAE)				Mean Absolute Percentage Error (MAPE)			
		in sample		out of sample		in sample		out of sample	
		mean	std	mean	std	mean	std	mean	std
M0	0.372	2491.1	9.6	2489.4	10.6	84.01%	0.26%	84.07%	0.44%
M1	0.896	968.9	3.6	969.8	3.5	27.52%	0.07%	27.54%	0.09%
M2	0.906	958.6	4.5	957.5	6.5	25.61%	0.07%	25.63%	0.09%

2.5 Optimization Model

We now formulate the Airline Network Planning with Supply and Demand inter-
actions (ANPSD) mathematically. The ANPSD optimizes network design and fleet
planning decisions for a single hub-and-spoke airline, with a profit-maximization ob-
jective.The ANPSD starts from a complete potential network, characterized by a set
of airport nodes ($\mathcal{N}$), including hub airports ($\mathcal{H}$) and spoke airports, and a set of can-
didate flight legs ($\mathcal{I}$). On the fleet side, the ANPSD considers a set of fleet types ($\mathcal{A}$)
along with aircraft specifications. Specifically, we consider cruising speed and average
landing and take-off times to compute flight block times (t_{ia}). This parameter is, in
turn, used to determine the feasibility of allocating aircraft type a to leg i, reflected
in the parameters δ_{ia}. We also define the seating capacity of each aircraft a (k_a).
Following Teodorović et al. (1994), we define an additional parameter of maximum
utilization (l_a) to account for maintenance and turnaround times.

To estimate operating profitability, the ANPSD needs inputs on airfares and op-
erating costs. Costs are divided into two components. Variable and semi-variable
costs include fuel, maintenance, flight personnel (pilots and cabin crew), airport and
air traffic control charges. These costs can be parametrized at the flight-aircraft level,
and are thus denoted by c_{ia}^{var}. Fixed costs capture aircraft ownership costs (depre-
ciation and/or leasing), and are denoted by c_a^{fix}. These fixed costs are critical to
appraise fleet changes accurately and support fleeting decisions. We omit general and
administrative overheads, which are not directly involved in strategic planning.

The key outputs of the model are well-defined network planning solutions, cap-
tured by the number of aircraft of each type (w_a) and the number of fights on each

Inputs: sets and parameters

$\mathcal{N}$	set of airports, indexed by n
$\mathcal{H} \subset \mathcal{N}$	subset of hub airports
$\mathcal{A}$	set of aircraft types, indexed by a
$\mathcal{I}$	set of flight legs, indexed by i or j
$\mathcal{I}_n^{out} \subset \mathcal{I}$	subset of outbound flight legs from airport $n \in \mathcal{N}$
$\mathcal{I}_n^{in} \subset \mathcal{I}$	subset of inbound flight legs to airport $n \in \mathcal{N}$
k_a	seat capacity of aircraft type a
l_a	maximum utilization of aircraft type a
p_i	average fare on flight leg i
t_{ia}	block time of leg i operated with aircraft type a
δ_{ia}	1 if flight leg i can be operated by fleet type a; 0 otherwise
δ_{in}^{out}	1 if flight leg i departs from airport n, i.e., $i \in \mathcal{I}_n^{out}$; 0 otherwise
δ_{in}^{in}	1 if flight leg i arrives at airport n, i.e., $i \in \mathcal{I}_n^{in}$; 0 otherwise
c_{ia}^{var}	trip cost for leg i operated by aircraft type a
c_a^{fix}	aircraft-related fixed cost of aircraft type a (e.g., ownership)
γ_i	leg-specific demand component (estimated from the demand model)
u, v	elasticity parameters (estimated from the demand model)
M	large parameter

Decision variables

$f_{ia} \in \mathbb{Z}^+$	frequency on flight leg i operated by aircraft a
$f_i \in \mathbb{Z}^+$	total frequency on flight leg i
$x_i \in \{0,1\}$	1 if flight leg i is operated; 0 otherwise
$q_i \in \mathbb{R}^+$	number of passengers accommodated on flight leg i
$s_i \in \mathbb{Z}^+$	number of spokes of flight legs i
$w_a \in \mathbb{Z}^+$	number of aircraft types a

leg by each aircraft type (f_{ia}). Through indirect variables, the model computes the supply-level inputs of the demand function (x_i, f_i and s_i), and replicates passenger flows (q_i).

Mathematical formulation

$$\max \quad \sum_{i \in I} p_i q_i - \sum_{i \in I} \sum_{a \in A} c_{ia}^{var} f_{ia} - \sum_{a \in A} c_a^{fix} w_a \qquad (2.2)$$

$$\text{s.t.} \quad \sum_{i \in I_n^{out}} f_{ia} = \sum_{i \in I_n^{in}} f_{ia} \qquad\qquad \forall n \in \mathcal{N}, \forall a \in \mathcal{A} \qquad (2.3)$$

$$\sum_{i \in I} f_{ia} t_{ia} \leq l_a w_a \qquad\qquad \forall a \in \mathcal{A} \qquad (2.4)$$

$$f_{ia} \leq \delta_{ia} f_{ia} \qquad\qquad \forall a \in \mathcal{A}, \forall i \in \mathcal{I} \qquad (2.5)$$

$$q_i \leq \sum_{a \in A} f_{ia} k_a \qquad\qquad \forall i \in \mathcal{I} \qquad (2.6)$$

$$f_i \leq \sum_{a \in A} f_{ia} \qquad\qquad \forall i \in \mathcal{I} \qquad (2.7)$$

$$f_i - x_i \geq 0 \qquad\qquad \forall i \in \mathcal{I} \qquad (2.8)$$

$$f_i - M x_i \leq 0 \qquad\qquad \forall i \in \mathcal{I} \qquad (2.9)$$

$$s_i = \sum_{n \in \mathcal{H}} \sum_{j \in I_n^{in}} \delta_{in}^{out} x_j + \sum_{n \in \mathcal{H}} \sum_{j \in I_n^{out}} \delta_{in}^{in} x_j \qquad \forall i \in \mathcal{I} \qquad (2.10)$$

$$q_i \leq \gamma_i f_i^u s_i^v \qquad\qquad \forall i \in \mathcal{I} \qquad (2.11)$$

$$f_{ia} \in \mathbb{Z}^+ \qquad\qquad \forall a \in \mathcal{A}, \forall i \in \mathcal{I} \qquad (2.12)$$

$$f_i \in \mathbb{Z}^+, \ x_i \in \{0,1\}, \ q_i \in \mathbb{R}^+, \ s_i \in \mathbb{Z}^+ \qquad \forall i \in \mathcal{I} \qquad (2.13)$$

$$w_a \in \mathbb{Z}^+ \qquad\qquad \forall a \in \mathcal{A} \qquad (2.14)$$

Equation (2.2) maximizes operating profits. The revenue component is given by the product of each leg's average price (p_i) and the number of accommodated passengers (q_i). Costs are expressed as the sum of fixed and variable costs. Equation (2.3) enforces flow balance at each node by setting the number of inbound flights equal to the number of outbound flights, for each aircraft type and at each airport. In a multi-hub setting, Equation (2.3) allows for circular routing, hence directional imbalance. If symmetry is required, we can replace Equation (2.3) by the following symmetry-inducing constraints: $f_{ia} = f_{-i,a} \ \forall i \in I, \ \forall a \in A$, where "$-i$" denotes the leg opposite to leg i (i.e., the leg from d to o). Equation (2.4) ensures that total aircraft utilization does not exceed the maximum possible utilization. Equation (2.5)

ensures the feasibility of each aircraft-leg combination. Equation (2.6) bounds the number of accommodated passengers on every flight leg by the allocated seat capacity. Next, Equation (2.7) computes the frequency on every flight leg. Equations (2.8) and (2.9) define the binary variables x_i from the frequency variables f_i, and Equation (2.10) computes the number of spokes served by the network. Using these variables, Equation (2.11) imposes the demand constraint, ensuring that the number of accommodated passengers does not exceed the estimated demand—obtained from the demand function outlined in Section 2.4. Note that the frequency on each flight leg and the number of spokes in the network are endogenous variables of the optimization model but all other predictors from Section 2.4 are exogenous. Therefore, the full demand specification (M2) can be succinctly represented as $D_i(f_i, s_i) = \gamma_i f_i^u s_i^v$, where u and v are the estimated elasticity parameters, and $\gamma_i > 0$ is a leg-specific parameter that includes the intercept and all remaining explanatory variables. Finally, Equations (2.12)–(2.14) define the domain of the variables.

ANPSD model structure

The ANPSD formulation results in a mixed integer non-linear, non-convex programming model. Integer variables are needed to represent the discrete decisions associated with flight frequencies, fleet composition, and route planning. The non-linearities and non-convexities arise from the demand function in Equation (2.11), needed to capture the demand-supply interactions. This demand function is equivalent to a Cobb-Douglas model based on leg frequency (f_i) and the number of spokes (s_i). Its Hessian is given by:

$$H(D_i) = \begin{pmatrix} \gamma_i u(u-1) f_i^{u-2} s_i^v & \gamma_i uv f_i^{u-1} s_i^{v-1} \\ \gamma_i uv f_i^{u-1} s_i^{v-1} & \gamma_i v(v-1) f_i^u s_i^{v-2} \end{pmatrix} \tag{2.15}$$

For D_i to be concave—hence, for the ASPD to be a convex optimization problem—we need:

$$\frac{\partial^2 D_i}{\partial f_i^2} \leq 0 \iff u(u-1) \leq 0$$

$$\frac{\partial^2 D_i}{\partial s_i^2} \leq 0 \iff v(v-1) \leq 0$$

$$\det H(D_i) \geq 0 \iff uv(1-u-v) \geq 0,$$

which yields (i) $0 \leq u \leq 1$, (ii) $0 \leq v \leq 1$, (iii) and $u + v \leq 1$. From Table 2.2, however, we obtained empirically $u = 0.9018$ and $v = 0.1514$, which satisfy conditions (i) and (ii) but not condition (iii). As a result, the two-dimensional demand function is not concave. From a computational standpoint, this structure makes ANPSD highly challenging to solve. Integer variables prevent the use of gradient-based algorithms for non-linear optimization (e.g., stochastic gradient descent). Non-linearities prevent the direct use of integer linear programming methods. And non-convexities prevent the use of simple gradient-based outer approximation schemes. Accordingly, in the next section, we present an original algorithm to solve the ANPSD.

2.6 Solution Algorithms

Mixed integer non-linear programming (MINLP) represents a broad class of optimization problems that combine the challenges of discrete and non-linear optimization. Typical solution methods can be classified into two broad categories. Branch-and-bound methods, first, follow a tree-based search strategy, by solving non-linear continuous subproblems iteratively and enforcing integrality via branching (e.g., Gupta and Ravindran 1985, Quesada and Grossmann 1992). Outer approximation methods, second, alternate between a relaxed master problem (which provides solution guarantees) and a subproblem that generates a feasible solution. Some of these outer approximation methods iterate between a mixed-integer master problem and non-linear continuous subproblems (e.g., Geoffrion 1972, Duran and Grossmann 1986).

Others only solve mixed-integer linear problems (of increasing size) by iteratively approximating the feasible region with first-order Taylor expansions (e.g., Westerlund and Pettersson 1995). Our algorithm falls into this latter category. These methods exhibit strong convergence properties for convex problems, collectively providing a comprehensive toolkit for convex MINLPs. In contrast, nonconvex MINLPs are much more challenging, and often solved by heuristics (e.g., Exler et al. 2008, Schlüter et al. 2009). General-purpose methods usually resort to discretization to approximate non-linear functions by means of piecewise linear functions. These methods induce a trade-off between computational times and solution quality, governed by coarse vs. granular discretization. Recent discretization formulations have been proposed to model non-linear and disjunctive constraints with a logarithmic number of variables and constraints, with substantial computational benefits (Vielma et al. 2010, Vielma and Nemhauser 2011). We use these methods as a benchmark in this paper.

In another stream of research, methods for convex MINLPs have been extended to broader classes of problems. For instance, Westerlund et al. (1998) and Still and Westerlund (2008) extend the extended cutting-plane (ECP) method from Westerlund and Pettersson (1995) to pseudo-convex and quasi-convex MINLPs. More recently, Eronen et al. (2015) extend it to non-smooth functions. These extensions correct the gradient-based linear approximation of a convex function by a factor α to account for non-convexities—this method is thus referred to as αECP. Results suggest that the extended cutting plane algorithm provides an effective approach for solving large-scale MINLP problems featuring a moderate degree of non-linearity. Our algorithm extends this body of work by generating two cutting planes at each iteration. We therefore refer to our algorithm as 2αECP.

2.6.1 2αECP: An Exact Subgradient-based Cutting Plane Algorithm

The proposed 2αECP algorithm develops iteratively a linear outer approximation of the non-linear, non-convex function (Equation (2.11)). If the demand function was

concave, the plane tangent to any feasible solution would provide a valid cut—akin to the ECP method. However, since the demand function is not concave, the tangent plane eliminates part of the feasible region. Instead, we demonstrate how *two semi-hyperplanes* can provide a valid outer approximation of the demand function. Accordingly, at each iteration, we solve a master problem by replacing Equation (2.11) by a set of outer-approximating semi-hyperplanes, and generate two new semi-hyperplanes. The algorithm updates a lower bound (feasible solution) and an upper bound (solution guarantee) at each iteration, ultimately converging to the optimal solution of ANPSD.

We now proceed to the description of this algorithm. All proofs are reported in Appendix 2.9.

Outer approximation with two semi-hyperplanes

For a given flight leg $i \in I$, we define the domain of D_i as $\Omega_i = \{(f_i, s_i) \in \mathbb{Z}^2 : 0 \leq f_i \leq \bar{f}, 0 \leq s_i \leq \bar{s}\}$ where $\bar{f}$ and $\bar{s}$ are appropriate upper bounds. Let $\nabla D_i(f_i, s_i) = (u\gamma_i f_i^{u-1} s_i^v, v\gamma_i f_i^u s_i^{v-1})$ be the gradient of D_i, defined over $\Omega_i \backslash \{(f_i, s_i) : f_i = 0 \vee s_i = 0\}$ (recall that $u < 1$ and $v < 1$). Consider $(f_i^\psi, s_i^\psi) \in \Omega_i$ such that $f_i^\psi \neq 0$ and $s_i^\psi \neq 0$ and let us define $\mathcal{W}_i^\psi(f_i, s_i)$ as follows:

$$\mathcal{W}_i^\psi(f_i, s_i) = \nabla D_i(f_i^\psi, s_i^\psi)^T \begin{pmatrix} f_i - f_i^\psi \\ s_i - s_i^\psi \end{pmatrix} = \gamma_i \left(f_i^\psi\right)^u \left(s_i^\psi\right)^v \left(\frac{u}{f_i^\psi} f_i + \frac{v}{s_i^\psi} s_i - u - v\right)$$

$$(2.16)$$

If the function D_i was concave, we would obtain an outer approximation as: $D_i(f_i^\psi, s_i^\psi) + \mathcal{W}_i^\psi(f_i, s_i) \geq D_i(f_i, s_i)$, $\forall (f_i, s_i) \in \Omega_i$. Instead, we search for a valid outer-approximation by partitioning the space Ω_i into $\Omega_{i1}, \cdots, \Omega_{iK}$ and searching for constants $\alpha_{i1}^\psi, \cdots, \alpha_{iK}^\psi$ such that:

$$D_i(f_i^\psi, s_i^\psi) + \alpha_{ik}^\psi \mathcal{W}_i^\psi(f_i, s_i) \geq D_i(f_i, s_i) \qquad \forall (f_i, s_i) \in \Omega_{ik} \qquad (2.17)$$

Proposition 1 provides such a valid approximation as the union of two semi-hyperplanes with an intersecting edge, shown in Figure 2-2. Specifically, we partition the feasible region based on whether $\mathcal{W}_i^\psi(f_i, s_i) \geq 0$ or $\mathcal{W}_i^\psi(f_i, s_i) < 0$, and define the following semi-hyperplanes:

$$\mathcal{L}_{i+}^\psi = \left\{ D_i(f_i^\psi, s_i^\psi) + \alpha_{i+}^\psi \mathcal{W}_i^\psi(f_i, s_i), (f_i, s_i) \in \Omega_{i+}^\psi \right\}$$
$$\mathcal{L}_{i-}^\psi = \left\{ D_i(f_i^\psi, s_i^\psi) + \alpha_{i-}^\psi \mathcal{W}_i^\psi(f_i, s_i), (f_i, s_i) \in \Omega_{i-}^\psi \right\}$$

The proof proceeds by construction, but needs to show that the function $D_i(f_i, s_i)$ achieves its maximum in (f_i^ψ, s_i^ψ) when $\mathcal{W}_i^\psi(f_i, s_i) = 0$, so the cuts remain valid where $\mathcal{W}_i^\psi(f_i, s_i) = 0$.

(a) $(f_i^\psi, s_i^\psi) = (300, 50)$ **(b)** $(f_i^\psi, s_i^\psi) = (300, 10)$
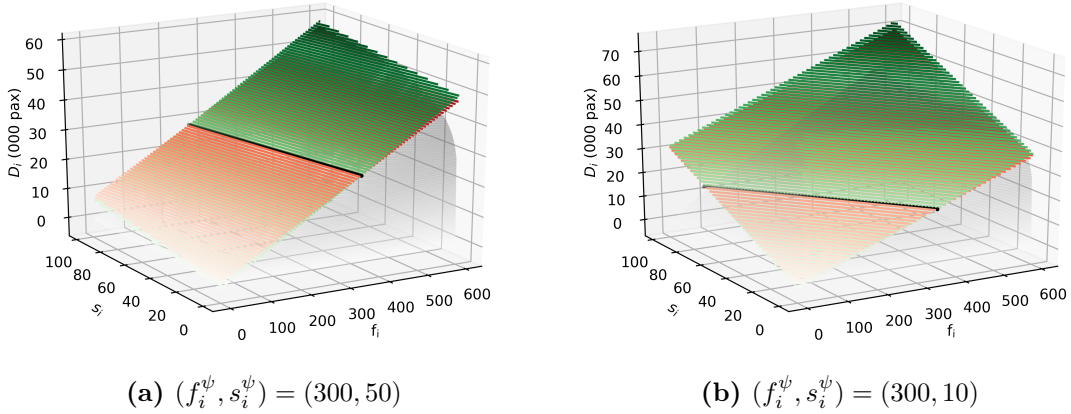
Figure 2-2: Illustration of valid cuts $\mathcal{L}_{i+}^\psi$ (green) and $\mathcal{L}_{i-}^\psi$ (red).

Proposition 1 *Let us partition Ω_i into $\Omega_{i+}^\psi \cup \Omega_{i-}^\psi$ as follows:*

$$\Omega_{i+}^\psi = \left\{ (f_i, s_i) \in \Omega_i : \frac{u}{f_i^\psi} f_i + \frac{v}{s_i^\psi} s_i - u - v \geq 0 \right\} \tag{2.18}$$

$$\Omega_{i-}^\psi = \left\{ (f_i, s_i) \in \Omega_i : \frac{u}{f_i^\psi} f_i + \frac{v}{s_i^\psi} s_i - u - v < 0 \right\} \tag{2.19}$$

There exist α_{i+}^{ψ} and α_{i-}^{ψ} such that:

$$D_i(f_i^{\psi}, s_i^{\psi}) + \alpha_{i+}^{\psi}\mathcal{W}_i^{\psi}(f_i, s_i) \geq D_i(f_i, s_i) \qquad \forall(f_i, s_i) \in \Omega_{i+}^{\psi} \tag{2.20}$$

$$D_i(f_i^{\psi}, s_i^{\psi}) + \alpha_{i-}^{\psi}\mathcal{W}_i^{\psi}(f_i, s_i) \geq D_i(f_i, s_i) \qquad \forall(f_i, s_i) \in \Omega_{i-}^{\psi} \tag{2.21}$$

We implement this outer approximation by means of the following linear relationships. In these equations, (f_i^{ψ}, s_i^{ψ}) denote linearization points; M_{i1}^{ψ}, M_{i2}^{ψ}, M_{i3}^{ψ} and M_{i4}^{ψ} denote "big-M" parameters; and δ_i denotes a binary variable equal to 1 if $(f_i, s_i) \in \Omega_{i+}^{\psi}$, and 0 otherwise.

$$\mathcal{W}_i^{\psi}(f_i, s_i) + M_{i1}^{\psi}(1 - \delta_i) \geq 0 \tag{2.22}$$

$$\mathcal{W}_i^{\psi}(f_i, s_i) - M_{i2}^{\psi}\delta_i \leq 0 \tag{2.23}$$

$$q_i \leq D_i(f_i^{\psi}, s_i^{\psi}) + \alpha_{i+}^{\psi}\mathcal{W}_i^{\psi}(f_i, s_i) + M_{i3}^{\psi}(1 - \delta_i) \tag{2.24}$$

$$q_i \leq D_i(f_i^{\psi}, s_i^{\psi}) + \alpha_{i-}^{\psi}\mathcal{W}_i^{\psi}(f_i, s_i) + M_{i4}^{\psi}\delta_i \tag{2.25}$$

Equations (2.22) and (2.23) enforce the logical relationship between the sign of $\mathcal{W}_i^{\psi}$ and δ_i. Note that these equations are linear because $\mathcal{W}_i^{\psi}$ is linear in f_i and s_i (Equation (2.16)). Equation (2.24) and (2.25) then select the valid (upper) plane accordingly. When $\mathcal{W}_i^{\psi}(f_i, s_i) \geq 0$ (and so $\delta_i = 1$) then Equation (2.24) is activated and Equation (2.25) is inactive—that is, the semi-hyperplane $\mathcal{L}_{i+}^{\psi}$ is activated. Vice versa, if $\mathcal{W}_i^{\psi}(f_i, s_i) < 0$ (and so $\delta_i = 0$), the semi-hyperplane $\mathcal{L}_{i-}^{\psi}$ is activated. To enforce these logical relationships with the tightest possible big-M parameters, we define:

$$M_{i1}^{\psi} = -\inf\left\{\mathcal{W}_i^{\psi}(f_i, s_i) : (f_i, s_i) \in \Omega_{i-}^{\psi}\right\} \tag{2.26}$$

$$M_{i2}^{\psi} = \sup\left\{\mathcal{W}_i^{\psi}(f_i, s_i) : (f_i, s_i) \in \Omega_{i+}^{\psi}\right\} \tag{2.27}$$

$$M_{i3}^{\psi} = \sup\left\{(\alpha_{i-}^{\psi} - \alpha_{i+}^{\psi})\mathcal{W}_i^{\psi}(f_i, s_i) : (f_i, s_i) \in \Omega_{i-}^{\psi}\right\} \tag{2.28}$$

$$M_{i4}^{\psi} = \sup\left\{(\alpha_{i+}^{\psi} - \alpha_{i-}^{\psi})\mathcal{W}_i^{\psi}(f_i, s_i) : (f_i, s_i) \in \Omega_{i+}^{\psi}\right\} \tag{2.29}$$

Equations (2.22)-(2.25) require computing the gradient of D_i in (f_i^{ψ}, s_i^{ψ}) and thus can only be used to build a valid outer approximation in differentiable linearization

points. However, the demand function $D_i(f_s, s_i)$ is not differentiable when $f_i = 0$ or $s_i = 0$. Proposition 2 defines valid outer approximations in those points where the demand function is not differentiable.

Proposition 2 *The following equations define valid outer approximations of $D_i(f_i, s_i)$ over $\{(f_i, s_i) : f_i = 0 \lor s_i = 0\}$:*

$$\text{for all } (f_i^\psi, s_i^\psi) \text{ such that } f_i^\psi \geq 0 \text{ and } s_i^\psi = 0: \quad q_i \leq D_i(\bar{f}, 1)s_i \tag{2.30}$$

$$\text{for all } (f_i^\psi, s_i^\psi) \text{ such that } f_i^\psi = 0 \text{ and } s_i^\psi > 0: \quad q_i \leq D_i(1, \bar{s})f_i \tag{2.31}$$

We define a master problem by replacing Equation (2.11) in the ANPSD by Equations (2.22)-(2.25), Equation (2.30) or Equation (2.31), for all flight legs and all linearization points $\{(f_i^\psi, s_i^\psi) \in \Omega_i, i \in \mathcal{I}, \psi \in \Psi\}$. This master problem is a mixed-integer linear problem. By construction, the master problem provides a relaxation of the ANPSD, hence an upper bound of the optimal solution.

Proposition 3 *Let $(f_{ia}^*, f_i^*, x_i^*, q_i^*, s_i^*, w_a^*)$ denote the optimal solution of the master problem. The objective function, given as follows, provides an upper bound to ANPSD.*

$$UB = \sum_{i \in I} p_i q_i^* - \sum_{i \in I} \sum_{a \in A} c_{ia}^{var} f_{ia}^* - \sum_{a \in A} c_a^{fix} w_a^*. \tag{2.32}$$

Next, from any solution of the master problem, we can retrieve a feasible solution of ANPSD by evaluating the true demand function and applying the aircraft capacity. In turn, we obtain a valid lower bound of the optimal solution. This is stated formally in Proposition 4.

Proposition 4 *Let $(f_{ia}^*, f_i^*, x_i^*, q_i^*, s_i^*, w_a^*)$ denote the optimal solution of the master problem. Then the following expression provides a valid lower bound to ANPSD:*

$$LB = \sum_{i \in I} p_i \cdot \min\left(\gamma_i (f_i^*)^u (s_i^*)^v, \sum_{a \in A} f_{ia}^* k_a\right) - \sum_{i \in I} \sum_{a \in A} c_{ia}^{var} f_{ia}^* - \sum_{a \in A} c_a^{fix} w_a^*. \tag{2.33}$$

Iterative algorithm

We now summarize the full iterative procedure in Algorithm 1. At each iteration, we solve the master problem and add a new set of cutting planes. This algorithm iteratively updates upper and lower bounds, thus providing a valid optimality gap at each iteration. The algorithm terminates when the optimality gap reaches a predefined tolerance ε.

Algorithm 1 2αECP algorithm.

Initialization:
- Iteration count $\psi = 0$, upper bound $UB^{(0)} = +\infty$, lower bound $LB^{(0)} = 0$
- Linearization points $(f_i^\psi, s_i^\psi) \in \Omega_i, \forall i \in \mathcal{I}$
- Master problem (Equations (2.2)–(2.10) and (2.12)-(2.14))

while $\frac{UB^\psi - LB^\psi}{LB^\psi} > \varepsilon$ **do**
    for $i \in \mathcal{I}$ **do**
        if $(f_i^\psi, s_i^\psi) \in \Omega_i \setminus \{(f_i, s_i) : f_i = 0 \vee s_i = 0\}$ **then**
            – compute $\mathcal{W}_i^\psi$ (Equation (2.16))
            – compute α_{i+}^ψ and α_{i-}^ψ (Equations (2.37)–(2.38))
            – compute M_{i1}^ψ, M_{i2}^ψ, M_{i3}^ψ and M_{i4}^ψ (Equations (2.26)–(2.29))
            – add cuts (2.22)-(2.25) to the master problem
        else
            – add cut (2.30) or (2.31) to the master problem
        end if
    end for
    – $\psi \leftarrow \psi + 1$
    – Solve the master problem
    – Store solution $(f_{ia}^\psi, f_i^\psi, x_i^\psi, q_i^\psi, s_i^\psi, w_a^\psi)$
    – Update the upper bound UB^ψ and the lower bound LB^ψ (Prop. 3-4)
end while
– Return $\left(f_{ia}^\psi, f_i^\psi, x_i^\psi, \min\left(\gamma_i(f_i^\psi)^u(s_i^\psi)^v, \sum_{a \in A} f_{ia}^\psi k_a \right), s_i^\psi, w_a^\psi \right)$

We now show that this algorithm is exact, that is, it converges in a finite number of iterations to the ANPSD optimum. Note that the upper bound is non-increasing—since we expand the set of cuts at each iteration—but the lower bound is not necessarily non-decreasing.

Proposition 5 *Algorithm 1 terminates in a finite number of iterations, and returns the optimal solution of ANPSD.*

2.6.2 Discretization Benchmarks

We evaluate the computational performances of 2αECP against two benchmarks. Both benchmarks rely on discretizing the space Ω_i and approximating the non-linear function $D_i(f_i, s_i)$ by a piecewise linear function. Under such discretization, the ANPSD can be reformulated as a mixed-integer linear program (MILP), which can be solved directly with commercial solvers.

Convex combination (CC)

This formulation directly approximates the non-linear function with a set of piecewise linear hyperplanes. Specifically, it considers a set of vertices in the three-dimensional space, of the form $\{(f_i^{(k)}, s_i^{(k)}, D_i(f_i^{(k)}, s_i^{(k)})) : k = 1, \cdots, K\}$. Each point $(f_i, s_i) \in \Omega_i$ lies in a triangle formed by three of these discretized points, say $(f_i^{(k_1)}, s_i^{(k_1)})$, $(f_i^{(k_2)}, s_i^{(k_2)})$, and $(f_i^{(k_3)}, s_i^{(k_3)})$. The CC method then expresses (f_i, s_i) as the convex combination of these three points, say $(f_i, s_i) = \lambda_1 \cdot (f_i^{(k_1)}, s_i^{(k_1)}) + \lambda_2 \cdot (f_i^{(k_2)}, s_i^{(k_2)}) + \lambda_3 \cdot (f_i^{(k_3)}, s_i^{(k_3)})$ (more generally, it expresses a p-dimensional vector as the convex combination of $p + 1$ adjacent points). The CC method proceeds by approximating the demand function as the corresponding convex combination of the demand values, i.e., $D_i(f_i, s_i) = \lambda_1 \cdot D_i(f_i^{(k_1)}, s_i^{(k_1)}) + \lambda_2 \cdot D_i(f_i^{(k_2)}, s_i^{(k_2)}) + \lambda_3 \cdot D_i(f_i^{(k_3)}, s_i^{(k_3)})$. We present the formulation for ANPSD in Appendix 2.9, and refer to Keha et al. (2004), Lee and Wilson (2001) and Padberg (2000) for more details.

Logarithmic branching convex combination (LOG)

This method follows the same general principle—approximating a non-linear function as the convex combination of discretized points. The key difference is that it reformulates the convex combination with a number of binary variables and constraints that grows logarithmically with the number of triangles (or polytopes), as opposed to linearly. Instead of adding one binary variable for each triangle, the LOG method encodes each triangle with a binary vector (of length the log-number of triangles) and implements a branching scheme based on SOS2 variables. As such, the

LOG method can yield significant computational benefits, especially for granular discretization schemes. We present the formulation for ANPSD in Appendix 2.9, and refer to Vielma et al. (2010) and Vielma and Nemhauser (2011) for details.

2.7   Computational Results

The performance and the scalability of our exact algorithm were tested through comprehensive computational experiments, using synthetic datasets inspired by real-world instances of varying size and complexity. We first present the experimental setup and the computational environment, and then compare the performance of our 2αECP against the discretization benchmarks.

2.7.1   Experimental Setup

The size of our instances is controlled by three parameters:

- number of hubs $|\mathcal{H}|$. We randomly select the hubs among the 10 largest hub airports in Europe.

- number of potential routes $|\mathcal{I}|$. We randomly select the potential routes by sampling spoke airports out of all commercial airports in Europe and connecting them to the hubs.

- maximum number of routes that can be operated, denoted by τ, enforced via the following additional constraint: $\sum_{i\in\mathcal{I}} x_i \leq \tau$. This constraint reflects airlines' network development practices, but also renders the problem more challenging computationally—thus enabling us to investigate the performance of the proposed solution algorithm.

Overall, we consider 9 scenarios, with 1 or 2 hubs, 100, 250 or 500 potential routes, and up to 100 or 250 actual routes. We run 5 instances for each scenario, leading to 45 instances in total. The problem sizes are consistent with the continental networks of major European airlines, such as KLM (one hub, 190 routes) and Air France (two hubs, 206 routes).

We consider five narrow-body aircraft types, representative of European continental fleets: A321 (200 seats), A320 (180 seats), A319 (144 seats), E190 (100 seats), and E175 (88 seats). In the absence of airline-specific data, we approximate operating costs following Swan and Adler (2006): $c_{ia}^{var} = (dist_i + 722) \cdot (k_a + 104) \cdot \$0.019 - c_a^{fix}/T_a$, where c_a^{fix} is the monthly leasing cost (0.9% of the market price) and T_a is the average number of trips per month (average utilization and block distance per aircraft type). We set the maximum daily utilization to 10 hours, in line with current practice. We define total fleet size according to real-world networks (70 aircraft for a 100-route network and 150 aircraft for a 250-route network). Finally, we compute price and demand parameters from the empirical relationships estimated in Section 2.4, and we estimate flight times as the ratio of the great circle distance between the origin and destination airports to aircraft cruising speed plus 30 minutes of climb and descent.

2.7.2   Computational Setup

The CC method is applied with 16 intervals on both dimensions, and the LOG method with 8, 16, or 32 intervals on both dimensions (yielding triangulations of 128, 512, and 2048 simplices, respectively). This setup enables us to investigate the trade-off between solution quality and computation time in the LOG discretization method.

We implement our 2αECP algorithm by solving the master problem at each iteration with a maximum runtime of 30 minutes and an optimality gap of 1% (except for the last iteration). We use a warm start with the solution from the previous iteration. Note that computing the α parameters (Equations (2.37)–(2.38)) and "big-M" parameters (Equations (2.26)–(2.29)) requires solving separate nonconvex MINLPs. However, given the finite solution space, their values can be computed efficiently by complete enumeration. We initialize the algorithm with an mid-point approximation—that is, $(\bar{f}/2, \bar{s}/2)$, for all flight legs, with $\bar{f} = 600$ (about 20 flights per day) and $\bar{s} = \tau/2$.

All models are implemented using Python 3.6 and the CPLEX MILP Solver (v12.9), on an Intel(R) Core(TM) i7-8700K CPU with a frequency of 3.70 GHz and 32 GB of RAM, and a maximum runtime of two hours.

For each scenario and each method, Table 2.4 reports the average number of variables (# cols) and constraints (# rows). Note that the logarithmic formulation reduces the number of variables by a factor of 2.6 and the number of constraints by 8.3, as compared to the CC method. Turning to our 2αECP algorithm, recall that the master problem increases in size at each iteration; accordingly, we report the problem size at the first and last iterations, showing the smallest and largest test instances. All problem instances in our 2αECP algorithm remain reasonable, comparable to LOG with $\Gamma = 8$ or $\Gamma = 16$ (obviously, this comes at the cost of solving multiple MILPs iteratively).

Table **2.4:** Model size.

| Network | CC | | | | LOG | | | | 2αECP | | | |
| | Γ = 16 | | Γ = 8 | | Γ = 16 | | Γ = 32 | | 1st iter | | last iter | |
$|\mathcal{H}| \times \tau \times |\mathcal{I}|$	# cols	# rows	# cols	# rows	# cols	# rows	# cols	# rows	# cols	# rows	# cols	# rows
1x100x100	80,358	30,665	9,628	3,286	30,460	3,682	110,019	4,079	1,006	1,912	1,890	5,448
1x100x250	201,210	76,768	24,101	8,209	76,265	9,203	275,482	10,196	2,486	4,724	4,149	11,376
1x100x500	403,386	153,894	48,312	16,446	152,892	18,438	552,288	20,430	4,986	9,474	7,406	19,154
1x250x500	403,386	153,894	48,312	16,446	152,892	18,438	552,288	20,430	4,986	9,474	8,868	25,002
2x100x100	81,011	30,917	9,711	3,317	30,711	3,717	110,911	4,117	1,011	1,917	1,976	5,777
2x100x250	200,567	76,525	24,028	8,188	76,024	9,178	274,599	10,169	2,481	4,710	4,341	12,150
2x100x500	402,095	153,405	48,162	16,398	152,406	18,384	550,519	20,369	4,981	9,460	8,062	21,784
2x250x250	200,567	76,525	24,028	8,188	76,024	9,178	274,599	10,169	2,491	4,729	4,344	12,141
2x250x500	403,391	153,899	48,317	16,451	152,897	18,443	552,293	20,435	5,001	9,498	7,791	20,658

2.7.3 Computational results

Table 2.5 reports the computational results. For both linearization benchmarks, we report the computation time (T, in seconds) and the percentage variation in operating profits compared to 2αECP ($\%\Delta_{SOL} > 0$ if the benchmark outperforms 2αECP, and $\%\Delta_{SOL} < 0$ otherwise)[3]. To ensure comparability, the solution of the linearization benchmarks, Z^*, refers to the objective value obtained by evaluating the profit function under the true demand function (as opposed to the discretized demand function). For our 2αECP algorithm, we report the objective value ($Z^{2\alpha ECP}$, in \$M) and the optimality gap achieved at the last iteration, after the two hours of computational time ($\%Gap$).

First, CC performs consistently poorly. Indeed, it results in worse solutions than the 2αECP algorithm in all 45 instances, with a relative difference ($\%\Delta_{SOL}$) of up to 98% in the worst case and of 55% on average. It also consistently underperforms LOG—and this would be exacerbated with coarser discretization (e.g., $\Gamma = 32$). Furthermore, the CC method uses the two-hour budget in most instances, so the poor solution quality is not outweighed by shorter runtimes. Ultimately, this shows the benefits of our algorithm, as compared to standard discretization.

Turning to the more advanced LOG discretization method, we observe a trade-off between solution quality and computational times, as expected. Using a coarse discretization with 8 intervals, the LOG method terminates fast in some instances. However, it results in significantly inferior solutions, as compared to the 2αECP algorithm—with a median difference of 5%. As the number of intervals increases to 16, the problem size increases accordingly (see Table 2.4), thus requiring longer runtimes. Yet, 2αECP still outperforms the linearization benchmark, resulting in better solutions in most instances, a median profit improvement of 1.3% and an average profit improvement of 3.5%. The main exception is scenario 1x100x500, under which LOG provides better solutions than 2αECP for all instances but one; however, for this instance (#4), LOG cannot even find a feasible solution within the

---

[3]$\%\Delta_{SOL} = (Z^* - Z^{2\alpha ECP})/Z^{2\alpha ECP}$, where Z^* and $Z^{2\alpha ECP}$ denote the values attained by CC/LOG and 2αECP.

Table 2.5: Results of computational experiments.

Network $|\mathcal{H}| \times \tau \times |\mathcal{I}|$	Inst.	CC $\Gamma = 16$ %Δ_{SOL}	T	LOG $\Gamma = 8$ %Δ_{SOL}	T	$\Gamma = 16$ %Δ_{SOL}	T	$\Gamma = 32$ %Δ_{SOL}	T	2αECP $Z^{2\alpha ECP}$	%Gap
1x100x100	#1	-84.9%	7200	-11.9%	7200	-4.1%	7200	-0.2%	7200	19.5	0.7%
	#2	-	7200	-8.6%	20	-3.9%	7200	-0.3%	7200	18.6	1.3%
	#3	-86.4%	7200	-6.2%	11	-2.0%	40	0.2%	410	21.6	1.0%
	#4	-88.1%	7200	-11.6%	14	-3.2%	116	0.0%	7200	18.6	0.9%
	#5	-75.4%	7200	-5.1%	32	-1.6%	7200	-0.5%	7200	21.4	0.8%
1x100x250	#1	-77.9%	7200	-2.7%	7200	-0.3%	7200	0.4%	7200	28.5	1.0%
	#2	-85.0%	7200	-2.9%	7200	0.4%	7200	0.5%	7200	29.4	1.1%
	#3	-84.9%	7200	-7.4%	7200	0.5%	7200	1.3%	7200	19.9	2.3%
	#4	-76.9%	7200	-4.7%	7200	-0.5%	7200	0.3%	7200	24.6	0.9%
	#5	-81.7%	7200	-2.4%	7200	0.6%	7200	0.5%	7200	29.2	1.5%
1x100x500	#1	-91.8%	7200	0.6%	7200	0.9%	7200	-	7200	35.8	2.0%
	#2	-82.0%	7200	0.4%	7200	1.6%	7200	0.9%	7200	35.2	2.5%
	#3	-81.4%	7200	1.5%	7200	2.0%	7200	-	7200	45.0	3.1%
	#4	-86.1%	7200	-2.2%	7200	-	7200	-	7200	35.0	3.9%
	#5	-86.5%	7200	0.5%	7200	2.3%	7200	-	7200	27.7	3.9%
1x250x500	#1	-0.9%	896	-4.9%	9	-0.9%	303	0.3%	4015	65.5	0.5%
	#2	-1.3%	7200	-10.0%	11	-1.3%	7200	0.3%	7200	41.1	0.5%
	#3	-88.7%	7200	-6.1%	11	-1.4%	972	0.2%	2389	63.9	0.4%
	#4	-1.0%	7200	-5.9%	12	-1.0%	273	0.6%	3013	61.1	1.0%
	#5	-1.4%	7200	-10.2%	10	-1.3%	7200	0.3%	6736	49.6	0.9%
2x100x100	#1	-3.2%	2937	-8.6%	11	-3.1%	241	0.8%	7200	13.6	2.8%
	#2	-3.0%	7200	-6.9%	15	-2.9%	285	-1.1%	7200	16.1	1.5%
	#3	-0.5%	7200	-2.8%	15	-0.5%	229	0.4%	7200	26.1	1.3%
	#4	-3.0%	6503	-6.6%	36	-2.9%	1821	0.3%	7200	17.3	1.7%
	#5	-39.8%	7200	-11.2%	32	-4.9%	719	-0.6%	7200	12.9	1.0%
2x100x250	#1	-78.4%	7200	-4.6%	7200	-2.8%	7200	-0.5%	7200	21.9	1.7%
	#2	-72.8%	7200	-5.7%	7200	-1.7%	7200	-0.2%	7200	22.6	1.1%
	#3	-1.3%	7200	-4.6%	7200	-1.3%	7200	0.2%	7200	34.0	0.8%
	#4	-79.7%	7200	-6.3%	7200	-2.1%	7200	-96.3%	7200	21.1	1.3%
	#5	-71.3%	7200	-4.7%	7200	-2.5%	7200	-0.4%	7200	21.6	0.8%
2x100x500	#1	-78.2%	7200	-2.6%	7200	1.0%	7200	-	7200	36.0	2.5%
	#2	-82.3%	7200	-3.6%	7200	-0.3%	7200	-	7200	29.6	2.9%
	#3	-82.7%	7200	-2.9%	7200	0.5%	7200	-	7200	30.7	2.5%
	#4	-70.3%	7200	-0.7%	7200	4.3%	7200	-	7200	31.7	4.9%
	#5	-1.6%	7200	-2.8%	7200	-0.4%	7200	-	7200	30.5	2.1%
2x250x250	#1	-82.0%	7200	-5.0%	1824	-7.4%	7200	-17.0%	7200	73.9	0.7%
	#2	-92.0%	7200	-5.7%	129	-1.3%	7200	-26.9%	7200	65.1	0.9%
	#3	-98.0%	7200	-5.2%	7200	-0.8%	7200	-	7200	54.7	3.3%
	#4	-86.0%	7200	-9.9%	7200	-6.5%	7200	-	7200	55.9	0.9%
	#5	-88.4%	7200	-4.4%	162	-0.7%	7200	-36.6%	7200	66.2	1.2%
2x250x500	#1	-1.8%	7200	-5.0%	7200	-1.5%	7200	0.0%	7200	60.2	3.4%
	#2	-1.2%	7200	-4.0%	7200	-0.5%	7200	0.2%	7200	77.2	1.8%
	#3	-0.7%	7200	-3.0%	67	-0.7%	7200	0.2%	7200	99.9	0.9%
	#4	-1.4%	3432	-5.9%	16	-1.3%	136	0.1%	6062	67.9	0.3%
	#5	-2.1%	3539	-6.3%	71	-2.0%	1306	0.0%	2705	52.9	0.4%
Summary	min	-98.0%	896	-11.9%	9	-100%	40	-100%	410	12.9	0.3%
	avg	-54.9%	6806	-5.1%	4065	-3.5%	5453	-28.3%	6664	38.5	1.6%
	max	-0.5%	7200	1.5%	7200	4.3%	7200	1.3%	7200	99.9	4.9%
	25th	-84.9%	7200	-6.3%	20	-2.5%	1821	-96.3%	7200	21.6	0.9%
	50th	-78.2%	7200	-5.0%	7200	-1.3%	7200	-0.2%	7200	30.7	1.2%
	75th	-2.1%	7200	-2.9%	7200	-0.3%	7200	0.3%	7200	54.7	2.3%

two-hour limit. This issue is exacerbated when we further increase the linearization granularity to 32 intervals. Indeed, the LOG method produces no feasible solution, or largely sub-optimal solutions, in 13 instances out of 45 (or over 25% of the cases). In contrast, 2αECP attains high-quality solutions with an optimality gap consistently below 5%. Furthermore, LOG requires the full two-hour computational budget in all but 7 instances, with 32 intervals—thus not markedly reducing runtimes as compared to the 2αECP algorithm.

Ultimately, the 2αECP algorithm yields consistenly good solutions, which considerably improve the quality of the solutions obtained with simple discretization schemes (e.g., the CC method). The 2αECP also dominates state-of-the-art discretization schemes (e.g., the LOG method) applied with coarse discretization, and also seems to improve the quality of the solutions obtained with highly granular discretization— resulting in better or comparable solutions in most instances, and achieving more consistent performance across all instances considered. Finally, unlike the discretization benchmarks, the 2αECP algorithm also provides solution quality guarantees, indicating a median optimality gap of 1.2% and a maximum optimality gap of 4.9% within the two-hour budget.

Finally, Figure 2-3 depicts the convergence pattern of the 2αECP per iteration (Figure 2-3a) and over time (Figure 2-3b). Each line represents the upper (blue) and lower (red) bounds in a single instance. Note that the number of iterations remains limited, ranging from 7 to 18. Moreover, most improvements are achieved during the first 1,000 seconds, thus highlighting the ability of 2αECP to compute good feasible solutions in short computation times.

This behavior can be explained as follows. In initial iterations, the algorithm roughly approximates the demand curves to narrow down the most profitable routes, resulting in strong improvements over the initial solution. In later iterations, in contrast, the algorithm refines capacity allocation decisions, with a more marginal impact on profitability. Moreover, the master problem becomes increasingly large from one iteration to the next, resulting in longer runtimes per iteration. These patterns can slow down convergence, especially when the number of allowed routes is small com-

(a) Convergence by number of iterations

(b) Convergence by time
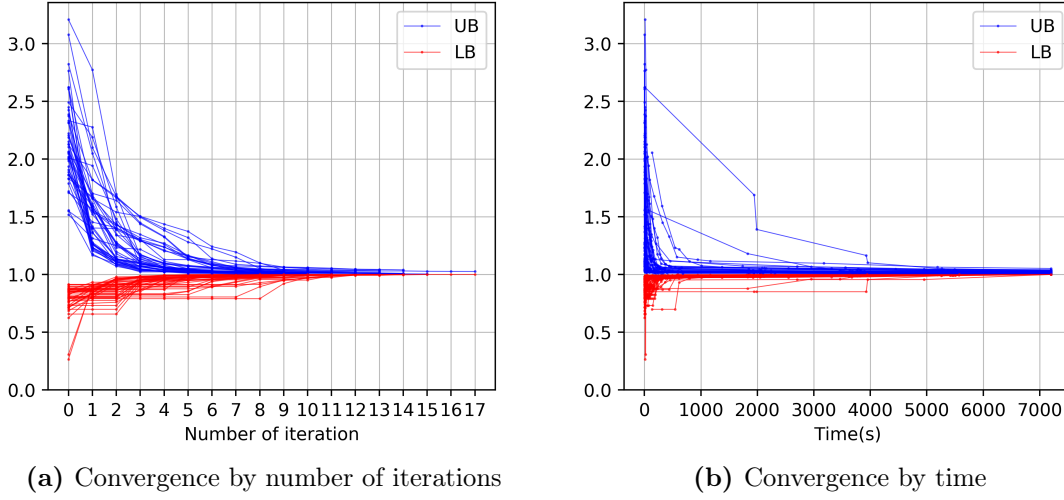
Figure 2-3: Convergence of 2αECP algorithm.

pared to the number of potential routes (e.g., 1x100x500, 2x100x500). In such cases, the algorithm takes more iterations to select the most profitable routes and tighten their outer demand approximation, since only a small subset of linearization points can be updated at each iteration. Yet, this issue does not significantly compromise the performance of 2αECP, indicated with the small optimality gaps achieved by the algorithm across all instances (Table 2.5 and Figure 2-3).

Overall, these computational experiments show that the 2αECP algorithm achieves stronger, more stable and more consistent performance than state-of-the-art benchmarks based on discretization and linearization. Moreover, 2αECP is an exact method that provides solution guarantees—hence, an optimality gap—at each iteration, as opposed to only generating a feasible solution.

2.8   Case Study

The application of the proposed modeling and algorithmic framework is now illustrated through a case study involving the continental network of Alitalia (in May 2018). Through this case study, we highlight the benefits of considering demand-supply interactions in airline network planning, and demonstrate how the ANPSD

can be used to support strategic decisions concerning the endogenous expansion (or contraction) of flight networks and fleets.

The continental network of Alitalia is based on a primary hub in Rome Fiumicino (FCO) and a secondary hub in Milan Linate (LIN). The network consists of 150 intra-European routes and the fleet is composed by 5 aircraft types: 6 Airbus A321, 28 A320, 19 A319, 5 Embraer E90, and 11 E75—a total of 69 aircraft. In the absence of proprietary data, we consider the same cost functions and parameters as in Section 2.7.1. Note that each airline has access to additional cost information, such as landing fees, costs of route openings, fleet costs, etc., which can easily be incorporated into the model. Despite these necessary simplifications, relative comparisons allow us to establish the benefits of the proposed approach.

To demonstrate the benefits of integrating demand-supply interactions in airline strategic planning, we solve the optimization model using the three demand specifications from Section 2.4:

– Demand specification M0, which ignores demand-supply interactions by considering inelastic demand estimates. In the optimization model, demand is treated as a fixed input parameter, replacing Equation (2.11) by a constraint of the form $q_i \leq \bar{D}_i$. The model is solved directly as a MILP. We refer to this model as $\mathcal{P}_0$, which replicates most of the literature.

– Demand specification M1, which partially captures the demand-supply interactions by modeling the effects of frequency, but disregards the positive effects generated by hub-and-spoke network effects. In the optimization model, Equation (2.11) is replaced by a constraint of the form $q_i \leq \gamma'_i f_i^{u'}$, which can be effectively approximated by (univariate) piecewise linearization methods (Vielma et al. 2010). We refer to this model as $\mathcal{P}_1$.

– Demand specification M2, which fully captures the demand-supply interactions by modeling the effects of frequency and network effects. This corresponds to the ANPSD (Section 2.5).

In this paper, we have considered specification M2 in the ANPSD formulation due

to its stronger statistical performance and theoretical grounding. At the same time, this choice has led to non-convexities, hence much more complex solution algorithms than with M0 and M1. The question is whether it was worth it—that is, whether the ANPSD outperforms $\mathcal{P}_0$ and $\mathcal{P}_1$.

We evaluate these three models under nine scenarios, by varying the number of routes (Δntw) and fleet size ($\Delta fleet$) by -20%/0%/+20% with respect to the baseline network. Table 2.6 reports two profit outcomes: (i) the objective function value of the model (Z, in \$M), and (ii) the profit obtained when evaluating each model's resulting solution with specification M2 (Z^*, in \$M). For instance, for $\mathcal{P}_0$, we obtain a network planning solution using M0, and we then evaluate the corresponding profits if the "true" demand is given by M0 (Z) or by the full demand specification M2 (Z^*). Comparing Z across different scenarios for for a given model identifies how the different models can support (or misguide) the evaluation of different network and fleet alternatives—using the relative variation of Z from the baseline scenario, denoted by $\%\Delta_{BL}$. On the other hand, Z^* identifies the "real" performance of each model, when evaluated against the best empirical specification—using the profit difference between $\mathcal{P}_0$ or $\mathcal{P}_1$ and ANPSD, denoted by Δ_{ANPSD}. We further report fixed costs, variable costs, fleet size, number of routes, passengers per flight and seats per flights.

Table 2.6: Performance of ANPSD with no, partial and full demand-supply interactions.

#	Scenario		Model	Z	$\%\Delta_{BL}$	Z^*	Δ_{ANPSD}	FixCost	VarCost	Fleet	Routes	Pax/flt	Seats/flt
	$\Delta fleet$	Δntw											
1)	0%	0%	$\mathcal{P}_0$	51.6	-	14.3	-40.42	17.8	8.8	19	144	180.7	181.2
	(baseline)		$\mathcal{P}_1$	114.0	-	6.7	-48.00	54.4	35.4	69	24	111.3	149.5
			$ANPSD$	54.7	-	54.7	-	54.4	32.9	69	150	79.5	152.0
2)	-20%	0%	$\mathcal{P}_0$	51.6	0.0%	14.3	-30.66	17.8	8.8	19	144	180.6	181.1
			$\mathcal{P}_1$	88.8	-22.1%	6.8	-38.15	41.8	25.4	53	20	106.8	150.7
			$ANPSD$	44.9	-17.8%	44.9	-	41.8	25.4	53	150	82.1	152.1
3)	+20%	0%	$\mathcal{P}_0$	51.5	0.0%	14.0	-47.70	18.0	8.6	19	144	181.6	181.9
			$\mathcal{P}_1$	132.7	16.4%	9.8	-51.97	63.8	41.7	81	28	112.5	149.5
			$ANPSD$	61.7	12.9%	61.7	-	63.8	38.9	81	150	78.9	152.1
4)	0%	-20%	$\mathcal{P}_0$	47.2	-8.4%	11.6	-40.04	16.4	7.6	17	120	185.9	186.2
			$\mathcal{P}_1$	114.0	0.0%	6.7	-44.97	54.4	35.4	69	24	111.3	149.5
			$ANPSD$	51.7	-5.6%	51.7	-	54.4	33.3	69	120	80.4	152.0
5)	0%	+20%	$\mathcal{P}_0$	67.0	30.0%	20.6	-43.62	22.4	11.1	24	174	179.3	179.6
			$\mathcal{P}_1$	118.8	4.2%	14.5	-49.76	54.4	33.0	69	26	103.9	150.6
			$ANPSD$	64.2	17.4%	64.2	-	54.4	34.0	69	180	86.6	151.8
6)	-20%	-20%	$\mathcal{P}_0$	47.2	-8.5%	12.0	-30.47	16.1	7.9	17	120	183.7	183.9
			$\mathcal{P}_1$	88.8	-22.1%	6.8	-35.65	41.8	25.4	53	20	106.8	150.7
			$ANPSD$	42.4	-22.4%	42.4	-	41.8	25.6	53	120	82.5	152.3
7)	+20%	-20%	$\mathcal{P}_0$	47.3	-8.3%	12.0	-46.13	16.4	7.3	17	120	181.0	181.1
			$\mathcal{P}_1$	132.7	16.4%	9.8	-48.33	63.8	41.7	81	28	112.5	149.5
			$ANPSD$	58.1	6.2%	58.1	-	63.8	39.2	81	120	79.1	152.4
8)	-20%	+20%	$\mathcal{P}_0$	67.0	30.0%	20.6	-31.86	22.3	11.2	24	174	178.5	178.8
			$\mathcal{P}_1$	91.9	-19.3%	7.7	-44.79	41.8	25.5	53	20	105.2	150.6
			$ANPSD$	52.5	-4.1%	52.5	-	41.8	26.0	53	180	88.9	151.7
9)	+20%	+20%	$\mathcal{P}_0$	67.0	30.0%	20.4	-52.25	22.6	11.0	24	174	180.0	180.5
			$\mathcal{P}_1$	139.0	22.0%	16.7	-55.94	63.8	39.2	81	30	104.5	150.6
			$ANPSD$	72.7	32.9%	72.7	-	63.8	40.1	81	180	85.7	152.0

$\%\Delta_{BL} = (Z_s - Z_{BL})/Z_{BL}$, where Z_s and Z_{BL} are, respectively, the profit achieved for scenario s (varying between 1 and 9) and for the baseline scenario. $\Delta_{ANPSD} = Z_s^* - Z_s^*(ANPSD)$, where Z_s is the profit achieved by the specific model in scenario s with demand specification $M2$, and $Z_s^*(ANPSD)$ is the corresponding profit achieved by the ANPSD model.

Let us start with our main observation: The ANPSD model considerably outperforms the two benchmarks, $\mathcal{P}_0$ and $\mathcal{P}_1$. Indeed, the ANPSD leads to profit outcomes ranging from 42.4\$M to 72.7\$M. In contrast, $\mathcal{P}_0$ and $\mathcal{P}_1$ lead to profit up to 20.6\$M and 16.7\$M, respectively, resulting in an average gap of 40.4 M\$ (72.5%) and 46.4 M\$ (83.4%). As compared to $\mathcal{P}_0$, ANPSD leverages demand stimulation mechanisms, capturing the impact of level of service on passenger demand. As compared to $\mathcal{P}_1$, ANPSD captures saturation patterns and accounts for the consolidation of hub-and-spoke operations. Perhaps surprisingly, $\mathcal{P}_1$ performs worse than $\mathcal{P}_0$, although demand specification M1 exhibits stronger empirical performance than M0. Overall, these results clearly demonstrate the benefits of fully capturing the demand-supply interactions (i.e., the effect of service frequency and of hub-and-spoke network effects on passenger demand) in airline strategic planning.

Let us now discuss the solutions of each model. Model $\mathcal{P}_0$ prioritizes the largest aircraft (i.e., A321 and A320) with a small number of flights. As a result, $\mathcal{P}_0$ leads to 180 seats per flight (as compared to 80 for ANPSD) but only 20 aircraft (as compared to 50–80 for ANPSD), thus maximizing load factors (in a somewhat unrealistic way). This is driven by the assumption that demand does not depend on the number of flights (frequency or number of spokes), so $\mathcal{P}_0$ takes a full supply-side perspective to match demand as cost-effectively as possible. Fleet variations thus do not impact estimated profits $\mathcal{P}_0$ ($\%\Delta_{BL} = 0\%$ in scenarios 2 and 3). In contrast, network expansion and contraction has a major effect on the $\mathcal{P}_0$ solution. Note that considering a maximum load factor as in (as Teodorović et al. 1994, Jaillet et al. 1996, for instance) would slightly increase the number of flights but would not significantly impact the qualitative insights.

Turning to $\mathcal{P}_1$, we observe the opposite distortions. Recall that our estimated coefficient for frequency under M1 is 1.06, indicative of increasing returns to scale. This leads to overconfident solutions in $\mathcal{P}_1$. Indeed, the model maximizes frequency on the routes with the highest contribution margins. As a result, the number of routes operated ranges from 20 to 30, less than 20% of all routes available. As such, network contraction does not impact the estimated profits from $\mathcal{P}_1$ ($\%\Delta_{BL} = 0\%$ in

scenario 4, and network expansion leads to small benefits (only two new routes added in scenario 5). In contrast, fleet variations have a very significant impact on the $\mathcal{P}_1$ solution.

The ANPSD achieves a middle ground between these two solutions, by balancing fleet and network considerations. When the network is fixed, a fleet contraction reduces profits by -17.8% (scenario 2), while a fleet expansion increases profits by +12.9% (scenario 3). This suggests that the current network is not saturated, yielding potential benefits by adding capacity on existing routes. When the fleet is fixed, a network contraction leads to a modest profit reduction by -5.6% (scenario 4). This is consistent with our observation from scenario 3, as the fleet capacity retrieved from the cancelled routes can be redeployed to strengthen supply on the most profitable routes. More importantly, in scenario 5, the model estimates a profit increase of 17.4% from network expansion. Obviously, the largest profit reduction (-22.4%) is attained in scenario 6, where both fleet and network are reduced, and the greatest profit increase (+32.9%) is expected under scenario 9. Note that, in scenario 9, the improvement is larger than for scenarios 3 and 5 taken together (+2.6%), which highlights the benefits of simultaneously optimizing fleeting and route selection.

Finally, scenarios 7 and 8 investigate intermediate situations in which the fleet is increased but the number of routes is decreased, and vice versa. Figure 2-4 illustrates the ANPSD outcome for these two scenarios, depicting network variations compared to the baseline. In scenario 7 (network contraction, fleet expansion), the model suggests dropping routes characterized by intense competition and lower prices, such as those from LIN to Southern Italy or Madrid (Figure 2-4a). As a result, operations are shrunk at LIN, refocusing on high-yield point-to-point markets towards major cities in Northern Europe. Accordingly, operations are more concentrated at the primary hub (FCO), leading to a hub consolidation strategy. In scenario 8 (fleet contraction, network expansion), in contrast, the model suggests two strategies (Figure 2-4b). First, it suggests opening new routes toward important destinations that are not currently served by Alitalia, such as Istanbul (IST), Lyon (LYS), Palma de Mallorca (PMI), Stuttgart (STR), and Wien (VIE). At the same time, the model suggests strategic

diversification of supply in the largest European multi-airport systems (e.g., London, Berlin, and Paris).
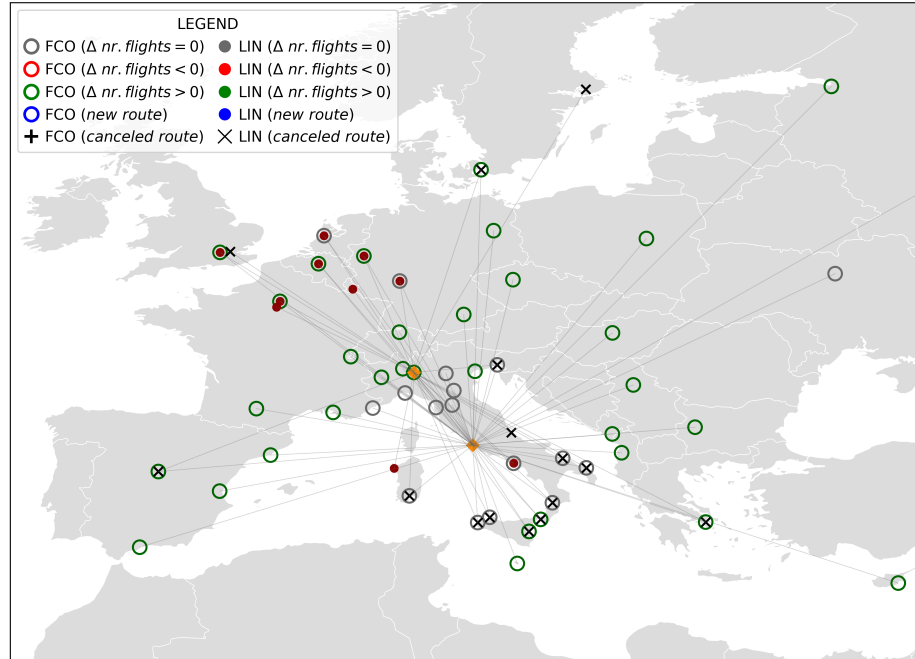
(a) $\Delta fleet = +20\% \ \Delta ntw = -20\%$
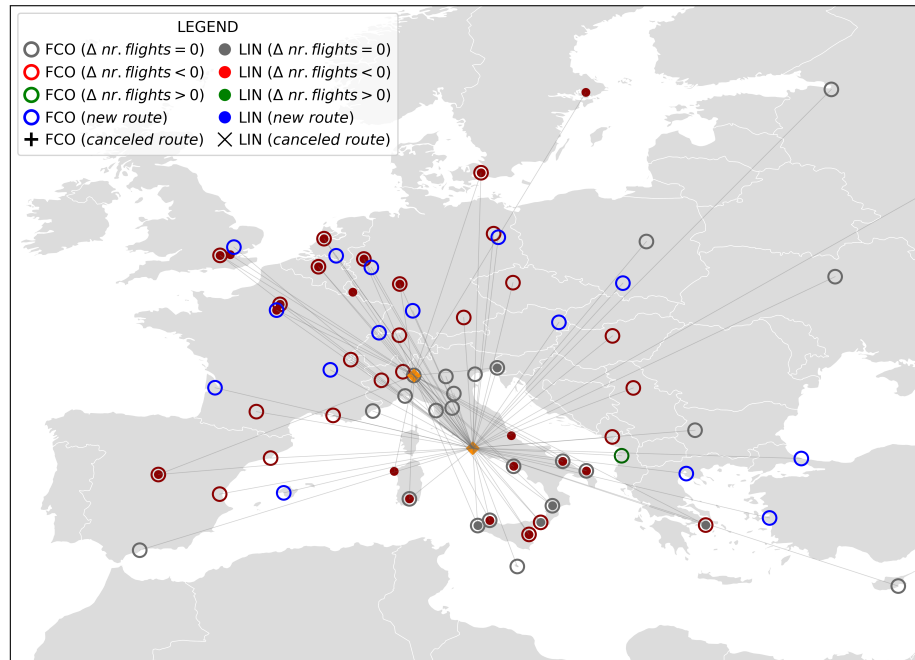
(b) $\Delta fleet = -20\% \ \Delta ntw = +20\%$

Figure 2-4: Impact of network and fleet expansion/contraction.

Overall, this analysis highlights the importance of jointly optimizing the different network planning decisions, spanning route planning, fight frequency and fleet composition.. The integrated ANPSD model can provide effective decision tools to support airline strategic planning. As the results have shown, the model's ability to fully capture the interactions between service frequency, network structure and passenger demand results in significant benefits, as compared to simpler approaches that ignore these demand-supply interactions

2.9 Conclusion

This paper has developed an original data-driven modeling and algorithmic framework to optimize airline network planning—referred to as Airline Network Planning with Supply and Demand interactions (ANPSD). The ANPSD simultaneously optimizes on which routes to operate, with which frequency and with which aircraft types, subject to operating constraints. The main novelty of ANPSD lies in the integration of an empirical demand model into a supply-side network optimization model, which enables an explicit treatment of demand-supply interactions. The model is formulated as a mixed integer non-convex optimization problem. To solve it, we have proposed an original exact cutting-plane algorithm, referred to as 2αECP, which generates iteratively an outer approximation of the non-linear, non-convex demand function by means of two hyperplanes.

Extensive computational experiments have shown the performance, scalability and practical applicability of the proposed modeling and computational approach. First, our exact 2αECP algorithm outperforms state-of-the-art benchmarks based on discretization and linearization. Second, the 2αECP algorithm yields solution quality guarantees—thus proving that it can generate near-optimal ANPSD solutions in reasonable computational times consistent with practical implementation requirements. Third, a case study using the continental network of Alitalia has shown that the ANPSD model provides much stronger planning solutions than baseline approaches based on inelastic demand estimates or partial treatments of demand-supply inter-

actions. Last, we showed how the ANPSD can practically aid decision-making by appraising different network and fleet expansion/contraction scenarios to inform the most promising development opportunities.

These positive results suggest future research opportunities in airline network planning. For instance, our aggregate leg-based demand model ignores heterogeneity across itineraries and origin-destination markets, which could be extended by developing data-driven estimates of passenger demand at a more disaggregate market level. Another simplification of ANPSD is the static treatment of inter-airline competition through its negative impact on prices. Extensions could incorporate game-theoretic considerations to systematically assess market dynamics among competing airlines. The ANPSD has also considered a deterministic demand model, which could be extended to capture demand uncertainty. Finally, the ANPSD has focused on maximizing operating profitability—the most important driver of network planning, but not the only one. An interesting avenue for future research lies in the in-depth consideration of additional factors underlying airline strategic planning, such as business models, the costs of opening or closing routes, and other practical considerations. The modeling and algorithmic framework provides methodological foundations to address these questions, with the ultimate potential of developing new data-driven approaches to support airline strategic planning that systematically integrate the interdependencies between demand and supply.

This paper comes at a critical time for the airline industry, in the midst of the COVID-19 pandemic. At a time of high uncertainty regarding the future of air travel, one thing is for sure: questions of airline network planning will be central in air transportation—for instance, Business Insider (2020) argued that "airlines face a years-long challenge to rebuild global route maps and networks that were devastated by the COVID-19 pandemic". By providing a new data-driven approach to support airline strategic planning, this paper strives to contribute to these efforts.

Appendix A: Proof of Statements

Proof of Proposition 1

Let us consider Ω_{i+}^{ψ} and Ω_{i-}^{ψ} given in Equations (2.18) and (2.19), and define Ω_{i0}^{ψ} as:

$$\Omega_{i0}^{\psi} = \left\{ (f_i, s_i) \in \Omega_i : \frac{u}{f_i^{\psi}} f_i + \frac{v}{s_i^{\psi}} s_i - u - v = 0 \right\} \tag{2.34}$$

Solving for α_{i+}^{ψ} and α_{i-}^{ψ} in Equations (2.20) and (2.21) yields:

$$\alpha_{i+}^{\psi} \geq \frac{D_i(f_i, s_i) - D_i(f_i^{\psi}, s_i^{\psi})}{\mathcal{W}_i^{\psi}(f_i, s_i)} \qquad \forall (f_i, s_i) \in \Omega_{i+}^{\psi} \setminus \Omega_{i0}^{\psi} \tag{2.35}$$

$$\alpha_{i-}^{\psi} \leq \frac{D_i(f_i, s_i) - D_i(f_i^{\psi}, s_i^{\psi})}{\mathcal{W}_i^{\psi}(f_i, s_i)} \qquad \forall (f_i, s_i) \in \Omega_{i-}^{\psi} \tag{2.36}$$

Since the feasible region is finite, we can define α_{i+}^{ψ} and α_{i-}^{ψ} as follows:

$$\alpha_{i+}^{\psi} = \sup \left\{ \frac{D_i(f_i, s_i) - D_i(f_i^{\psi}, s_i^{\psi})}{\mathcal{W}_i^{\psi}(f_i, s_i)} : (f_i, s_i) \in \Omega_{i+}^{\psi} \setminus \Omega_{i0}^{\psi} \right\} \tag{2.37}$$

$$\alpha_{i-}^{\psi} = \inf \left\{ \frac{D_i(f_i, s_i) - D_i(f_i^{\psi}, s_i^{\psi})}{\mathcal{W}_i^{\psi}(f_i, s_i)} : (f_i, s_i) \in \Omega_{i-}^{\psi} \right\} \tag{2.38}$$

To complete the proof, we need to show that Equation (2.20) holds for $(f_i, s_i) \in \Omega_{i0}^{\psi}$. To this end, we need to prove that $D_i(f_i^{\psi}, s_i^{\psi}) \geq D_i(f_i, s_i)$ for all $(f_i, s_i) \in \Omega_{i0}^{\psi}$, since $\mathcal{W}_i^{\psi}(f_i, s_i) = 0$ for all $(f_i, s_i) \in \Omega_{i0}^{\psi}$. In other words, we need to prove that $D_i(f_i, s_i)$ achieves its maximum over Ω_{i0}^{ψ} in (f_i^{ψ}, s_i^{ψ}). By solving for f_i in Equation (2.34), we define a function expressing the demand as a function of s_i (shown in Figure 2-5):

$$\text{For all } s_i \in \mathbb{Z} \cap \left[0, \frac{u+v}{v} s_i^{\psi} \right], \quad \phi_i^{\psi}(s_i) = D_i \left(\left(u + v - \frac{v}{s_i^{\psi}} s_i \right) \frac{f_i^{\psi}}{u}, s_i \right)$$

$$\phi_i^{\psi}(s_i) = \gamma_i \left[\left(u + v - \frac{v}{s_i^{\psi}} s_i \right) \frac{f_i^{\psi}}{u} \right]^u s_i^v$$
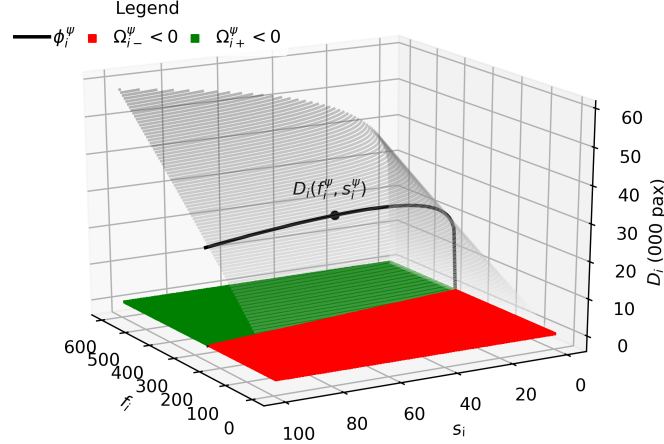
Figure 2-5: Illustration of ϕ_i^ψ (black solid line) for $(f_i^\psi, s_i^\psi) = (300, 50)$.

Note that D_i is differentiable over $\Omega_i \setminus \{(f_i, s_i) : f_i = 0 \vee s_i = 0\}$. Therefore, ψ is differentiable over $\left(0, \frac{u+v}{v} s_i^\psi\right)$. Since $\phi_i^\psi(0) = 0$, the maximum is not attained in 0. We have, for all $s_i \left(0, \frac{u+v}{v} s_i^\psi\right)$:

$$
\left(\phi_i^\psi\right)'(s_i) = \gamma_i \left[v s_i^{v-1} \left[\left(u + v - \frac{v}{s_i^\psi} s_i\right) \frac{f_i^\psi}{u}\right]^u - \frac{v}{s_i^\psi} \frac{f_i^\psi}{u} s_i \cdot u \left[\left(u + v - \frac{v}{s_i^\psi} s_i\right) \frac{f_i^\psi}{u}\right]^{u-1}\right]
$$

$$
= \gamma_i \left(\frac{f_i^\psi}{u}\right)^u \left(\frac{v}{s_i^\psi}\right)^{u-1} \left(\frac{u+v}{v} - s_i\right)^{u-1} s_i^{v-1} \cdot v \cdot (u+1) \left(1 - \frac{s_i}{s_i^\psi}\right)
$$

Therefore, ϕ_i^ψ is increasing over $\left[0, s_i^\psi\right)$ and decreasing over $\left(s_i^\psi, \frac{u+v}{v} s_i^\psi\right]$. We obtain that s_i^ψ is a global maximum of ϕ_i^ψ, hence (f_i^ψ, s_i^ψ) is a global maximum of D_i over Ω_{i0}^ψ. ∎

Proof of Proposition 2

We need to prove that $D_i(\bar{f}, 1) s_i \geq D_i(f_i, s_i)$, $\forall (f_i, s_i) \in \Omega_i$. First, note that D_i is concave with respect to s_i. Indeed, the second-order partial derivative of D_i with respect to s_i is given by $\partial^2 D_i / \partial s_i^2 = \gamma_i v(v-1) f_i^u s_i^{v-2} < 0$. Along with the integrality of s_i, this implies that: $D_i(f_i, 1) s_i \geq D_i(f_i, s_i)$, $\forall (f_i, s_i) \in \Omega_i$. Moreover, D_i is increasing with respect to f_i. Indeed, $\partial D_i / \partial f_i = \gamma_i u f_i^{u-1} s_i^v > 0$. There-

fore, $D(\bar{f}, s_i) \geq D(f_i, s_i) \; \forall (f_i, s_i) \in \Omega_i$. We thus obtain $D_i(\bar{f}, 1)s_i \geq D_i(\bar{f}, s_i) \geq D(f_i, s_i), \; \forall (f_i, s_i) \in \Omega_i$. This completes the proof of Equation (2.30). Equation (2.31) follows similarly. ∎

Proof of Proposition 3

This follows directly from Proposition 1 and Proposition 2, which show that Equations (2.22)-(2.25), Equation (2.30) and Equation (2.31) provide valid outer approximation of Equation (2.11). Thus, the master problem is a relaxation of ANPSD, and its optimal solution is an upper bound of the ANPSD optimum. ∎

Proof of Proposition 4

The proof relies on three observations. First, the constraints of the master problem and the ANPSD are identical, except Equation (2.11). Second, Equation (2.11) involves only the decision variables q_i. Third, the decision variables q_i are only involved in Equations (2.6) and (2.11). Therefore, if $q'_i = \min\left(\gamma_i(f_i^*)^u(s_i^*)^v, \sum_{a \in A} f_{ia}^* k_a\right)$, then $(f_{ia}^*, f_i^*, x_i^*, q'_i, s_i^*, w_a^*)$ is a feasible solution of the ANPSD, which achieves the objective given in Equation (2.33). The proof follows. ∎

Proof of Proposition 5

First, if the solution obtained by the master problem at iteration ψ is a feasible solution for ANPSD, than this point is a global optimum for ANPSD. Indeed, let $\mathcal{Z}^\psi$ denote the solution value of the master problem at iteration ψ and let $\mathcal{Z}^{opt}$ represent the optimal solution of ANPSD. From proposition 3 we know that $\mathcal{Z}^\psi \geq \mathcal{Z}^{opt}$. But if $(f_{ia}^\psi, f_i^\psi, x_i^\psi, q_i^\psi, s_i^\psi, w_a^\psi)$ is feasible under ANPSD, it follows that $\mathcal{Z}^\psi \leq \mathcal{Z}^{opt}$. Hence, we obtain $\mathcal{Z}^\psi = \mathcal{Z}^{opt}$. Therefore, the algorithm terminates at an optimal solution of ANPSD.

Let us now consider the case in which the solution attained by the master problem at iteration ψ is infeasible for ANPSD. Since the master problem only relaxes the nonlinear constraint (Equation (2.11)), we know that there exists $i \in \mathcal{I}$ such that

$q_i^\psi > \gamma_i (f_i^\psi)^u (s_i^\psi)^v$ (otherwise, the solution would be feasible). In the next iteration, a new outer approximation is added at each (f_i^ψ, s_i^ψ) according to Equations (2.22)-(2.25), Equation (2.30) or Equation (2.31)). To this end, we make three observations:

1. By construction, the new cutting planes make the previous solution infeasible—so any previously visited sub-optimal solution will never be revisited by the master problem, that is:

$$(f_{ia}^\psi, f_i^\psi, x_i^\psi, q_i^\psi, s_i^\psi, w_a^\psi) \neq (f_{ia}^\eta, f_i^\eta, x_i^\eta, q_i^\eta, s_i^\eta, w_a^\eta), \ \forall \eta < \psi,$$

2. The cutting planes bound the variables q_i to the true demand values in (f_i^ψ, s_i^ψ), since $(f_i^\psi, s_i^\psi) \in \Omega_{i0}^\psi$ (Proposition 1). Therefore, a solution $(f_{ia}^\psi, f_i^\psi, x_i^\psi, q_i^\psi, s_i^\psi, w_a^\psi)$ is a feasible ANPSD solution if there exists $\eta < \psi$ such that, for all $i \in \mathcal{I}$, $(f_i^\psi, s_i^\psi) = (f_i^\eta, s_i^\eta)$.

3. The set of feasible values of f_i and s_i in ANPSD is finite.

We conclude that the algorithm has finite convergence. Indeed, assume by contradiction that the series of master problem solutions does not converge in a finite number of iterations. Then, it would generate an infinite series of infeasible points for ANPSD, each different from one another (See 1. above) and each featuring (at least) one new value of f_i or s_i (See 2. above), which contradicts the fact that the set of points (f_i, s_i) is finite (See 3. above). This completes the proof. ∎

Appendix B: Formulation of the Discretization Benchmarks

Convex combination (CC).

Let Γ_f and Γ_s be two sets of intervals partitioning $[0, \bar{f}]$ and $[0, \bar{s}]$, respectively. Let $\mathcal{B}_f$ and $\mathcal{B}_s$ be the corresponding sets of breakpoints. This discretization partitions Ω_i into a grid of rectangles. We further divide each rectangle into two triangles to partition the two-dimensional domain into a set of mutually exclusive triangles $\mathcal{T}$ and define $\mathcal{T}_{b_f b_s} = \{t \in \mathcal{T} : (b_f, b_s) \in \mathcal{T}\}$. The CC formulation represents each point (f_i, s_i) as a convex combination of the three vertices of the triangle that contains it. Specifically, let us introduce $\lambda^i_{b_f b_s}$ as a positive variable defined over the set of vertices $(b_f, b_s) \in \mathcal{B}_f \times \mathcal{B}_s$ and representing the weights in the convex combination. We further define ζ^i_t as a binary variable equal to 1 if triangle $t \in \mathcal{T}_{b_f b_s}$ is activated, and zero otherwise. We then replace the nonlinear constraint (Equation (2.11)) with the following linear constraints.

$$f_i = \sum_{b_f \in \mathcal{B}_f} \sum_{b_s \in \mathcal{B}_s} \lambda^i_{b_f b_s} b_f \qquad \forall i \in \mathcal{I} \qquad (2.39)$$

$$s_i = \sum_{b_f \in \mathcal{B}_f} \sum_{b_s \in \mathcal{B}_s} \lambda^i_{b_f b_s} b_s \qquad \forall i \in \mathcal{I} \qquad (2.40)$$

$$q_i \leq \sum_{b_f \in \mathcal{B}_f} \sum_{b_s \in \mathcal{B}_s} \lambda^i_{b_f b_s} \gamma_i (b_f)^u (b_s)^v \qquad \forall i \in \mathcal{I} \qquad (2.41)$$

$$\sum_{b_f \in \mathcal{B}_f} \sum_{b_s \in \mathcal{B}_s} \lambda^i_{b_f b_s} = 1 \qquad \forall i \in \mathcal{I} \qquad (2.42)$$

$$\lambda^i_{b_f b_s} \leq \sum_{t \in \mathcal{T}_{b_f b_s}} \zeta^i_t \qquad \forall b_f \in \mathcal{B}_f, \forall b_s \in \mathcal{B}_s, \forall i \in \mathcal{I} \qquad (2.43)$$

$$\sum_{t \in \mathcal{T}} \zeta^i_t = 1 \qquad \forall i \in \mathcal{I} \qquad (2.44)$$

$$\lambda^i_{b_f b_s} \geq 0 \qquad \forall b_f \in \mathcal{B}_f, \forall b_s \in \mathcal{B}_s, \forall i \in \mathcal{I} \qquad (2.45)$$

$$\zeta^i_t \in \{0, 1\} \qquad \forall t \in \mathcal{T}, \forall i \in \mathcal{I} \qquad (2.46)$$

Logarithmic branching convex combination (LOG).

Let Γ_f and Γ_s be two sets of intervals partitioning $[0, \bar{f}]$ and $[0, \bar{s}]$, respectively, such that $|\Gamma_f|$ and $|\Gamma_f|$ are powers of two. Let $\mathcal{B}_f$ and $\mathcal{B}_s$ be the corresponding breakpoints. This discretization again partitions Ω_i into a grid of rectangles. Each rectangle is then divided into two triangles, using Union Jack triangulation. This partition strategy allows an effective branching scheme that first selects a rectangle by applying SOS2 branching on the two dimensions separately, and then selects one of the two triangles inside the rectangle by forbidding one of the two non-shared vertices.

We define a SOS2 compatible function Λ to map each set of intervals into binary vectors of length $\log_2(|\Gamma_f|)$ and $\log_2(|\Gamma_s|)$, that is, $\Lambda : \Gamma_f \rightarrow \{0, 1\}^{\log_2(|\Gamma_f|)}$ and $\Lambda : \Gamma_s \rightarrow \{0, 1\}^{\log_2(|\Gamma_s|)}$. SOS2 compatible functions are such that adjacent intervals only differ by one element, according to the gray code property. We label as $l_f = 1, \cdots, \log_2(|\Gamma_f|)$ (resp. $l_s = 1, \cdots, \log_2(|\Gamma_s|)$) the elements of the binary codes. We define $\mathcal{B}_f^+(l_f, \Lambda)$, $\mathcal{B}_f^0(l_f, \Lambda)$, $\mathcal{B}_s^+(l_s, \Lambda)$ and $\mathcal{B}_s^0(l_s, \Lambda)$ to represent the subsets of breakpoints whose adjacent intervals have the l_f binary digit equal to 1 or 0. We further define $\mathcal{B}_f^{even}$ and $\mathcal{B}_f^{odd}$ (resp. $\mathcal{B}_s^{even}$ and $\mathcal{B}_s^{odd}$) to include even and odd breakpoints, respectively.

We introduce auxiliary SOS2 variables $\lambda_{b_f b_s}^i$ defined over the set of vertices $(b_f, b_s) \in \mathcal{B}_f \times \mathcal{B}_s$ and representing the weights in the convex combination. We also introduce auxiliary binary variables $\eta_{l_f}^i$, $\eta_{l_s}^i$ and ξ_i to implement the branching scheme.

We then replace the nonlinear constraint (Equation (2.11)) with the following linear constraints.

$$f_i = \sum_{b_f \in \mathcal{B}_f} \sum_{b_s \in \mathcal{B}_s} \lambda^i_{b_f b_s} b_f \qquad \forall i \in \mathcal{I} \tag{2.47}$$

$$s_i = \sum_{b_f \in \mathcal{B}_f} \sum_{b_s \in \mathcal{B}_s} \lambda^i_{b_f b_s} b_s \qquad \forall i \in \mathcal{I} \tag{2.48}$$

$$q_i \leq \sum_{b_f \in \mathcal{B}_f} \sum_{b_s \in \mathcal{B}_s} \lambda^i_{b_f b_s} \gamma_i (b_f)^u (b_s)^v \qquad \forall i \in \mathcal{I} \tag{2.49}$$

$$\sum_{b_f \in \mathcal{B}_f} \sum_{b_s \in \mathcal{B}_s} \lambda^i_{b_f b_s} = 1 \qquad \forall i \in \mathcal{I} \tag{2.50}$$

$$\sum_{b_f \in \mathcal{B}_f^+(l_f, \Lambda)} \sum_{b_s \in \mathcal{B}_s} \lambda^i_{b_f b_s} \leq \eta^i_{l_f} \qquad \forall l_f = 1, \cdots, \log_2(|\Gamma_f|), \forall i \in \mathcal{I} \tag{2.51}$$

$$\sum_{b_f \in \mathcal{B}_f^0(l_f, \Lambda)} \sum_{b_s \in \mathcal{B}_s} \lambda^i_{b_f b_s} \leq 1 - \eta^i_{l_f} \qquad \forall l_f = 1, \cdots, \log_2(|\Gamma_f|), \forall i \in \mathcal{I} \tag{2.52}$$

$$\sum_{b_f \in \mathcal{B}_f} \sum_{b_s \in \mathcal{B}_s^+(l_s, \Lambda)} \lambda^i_{b_f b_s} \leq \eta^i_{l_s} \qquad \forall l_s = 1, \cdots, \log_2(|\Gamma_s|), \forall i \in \mathcal{I} \tag{2.53}$$

$$\sum_{b_f \in \mathcal{B}_f} \sum_{b_s \in \mathcal{B}_s^0(l_s, \Lambda)} \lambda^i_{b_f b_s} \leq 1 - \eta^i_{l_s} \qquad \forall l_s = 1, \cdots, \log_2(|\Gamma_s|), \forall i \in \mathcal{I} \tag{2.54}$$

$$\sum_{b_f \in \mathcal{B}_f^{even}} \sum_{b_s \in \mathcal{B}_s^{odd}} \lambda^i_{b_f b_s} \leq \xi_i \qquad \forall i \in \mathcal{I} \tag{2.55}$$

$$\sum_{b_f \in \mathcal{B}_f^{odd}} \sum_{b_s \in \mathcal{B}_s^{even}} \lambda^i_{b_f b_s} \leq 1 - \xi_i \qquad \forall i \in \mathcal{I} \tag{2.56}$$

$$\lambda^i_{b_f b_s} \geq 0 \qquad \forall b_f \in \mathcal{B}_f, \forall b_s \in \mathcal{B}_s, \forall i \in \mathcal{I} \tag{2.57}$$

$$\eta^i_{l_f}, \eta^i_{l_s} \in \{0, 1\} \qquad \forall l_f = 1, \cdots, \log_2(|\Gamma_f|), \forall l_s = 1, \cdots, \log_2(|\Gamma_s|) \tag{2.58}$$

$$\xi_i \in \{0, 1\} \qquad \forall i \in \mathcal{I} \tag{2.59}$$

Chapter 3

Integrated Flight Scheduling and Fleet Assignment with Improved Supply-demand Interactions

3.1 Introduction

The airline industry is a major economic force in today's global economy, not only for its direct economic impact but also for its great multiplicative effect on international business and tourism. Alongside the capital-intensive nature and large-scale operations, the many challenges faced by airlines to keep up with the increasingly competitive and fast-changing environment make the development of effective planning processes of outmost importance. In this regard, grounding on accurate forecasts of air travel demand is crucial. Starting from the longest-term strategic decisions to daily schedules, relying on proper traffic estimates is key to make optimal decisions at any stage of the planning process, and eventually deliver airlines' success.

Traditionally, an airline planning process goes through several sequential steps (Belobaba et al. 2015, Dong et al. 2016, Jamili 2017). Once the first phase—i.e. route development—has been addressed and potential city-pairs to serve have been decided, the next two steps are flight scheduling and fleet assignment. Flight schedul-

ing involves determining the itineraries to be operated in terms of frequency and timetable (Vaze and Barnhart 2012), while fleet assignment is about assigning appropriate fleet types to each flight leg such that seat capacity optimally matches the expected demand, subject to resource balance constraints in the network. Following the fleet assignment, aircraft rotation and crew scheduling problems are addressed (see, e.g., Barnhart et al. (2003b) and Gronkvist (2005), respectively, for a review).

Owing to the complexity and large scale of many real-world airline operations, solving the entire airline planning process simultaneously is unfeasible and not necessarily desirable in light of the different timeframes and levels of detail required by the respective planning steps. Accordingly, separate models have been proposed for the respective sub-problems; however, although such models are more practically solvable, they do not fully capture the interdependencies across planning decisions. One of the most complex interactions to model is the relationship between demand and supply. To tackle this issue, researchers have been developing integrated flight scheduling and fleet assignment models in which demand-supply interactions are explicitly taken into account.

The motivation for this research comes from the fact that current approaches focus primarily on the redistribution of passenger flows among competing itineraries (demand allocation) but tend to overlook the potential demand stimulation effects that result from improving the air service provision in city-pair markets (demand generation), or treat them using simplified aggregate models. As a result, such models are likely to work well in situations in which schedule changes are limited and markets are dense, while potentially failing to assess realistic development opportunities in thinner markets or in general situations in which the demand is significantly induced rather than captured from the existing competition.

This paper's main contribution is the integration of a hierarchical demand model based on a nested logit formulation into a flight scheduling and fleet assignment model, which results in a new flexible, utility-based nonlinear mixed-integer formulation that consistently incorporates the combined effect of demand generation and allocation on individual itineraries managed by a given airline. The solvability of the

integrated model for mid-sized hub-and-spoke networks is demonstrated by means of a comprehensive computational study and the managerial insights that can be derived from its application are illustrated through a case study involving the domestic and international network of the Italian flag carrier Alitalia based at Milan Linate airport.

The remainder of this paper is organized as follows. In Section 3.2, we explore the relevant literature and clarify the contribution of this work. Section 3.3 describes the model formulation, in which we first present the demand model and then discuss how to embed it effectively into the optimization model. Section 3.4 provides the results of the computational experiments, and Section 3.5 describes the application of the model to the Alitalia network. Finally, in Section 3.6 we conclude the paper and discuss directions for future research.

3.2 Literature Review

3.2.1 Integrated flight scheduling and fleet assignment models

Flight scheduling and fleet assignment have received increasing attention in the optimization literature owing to their high degree of impact on airline profits. Models of increasing complexity have been developed to embrace the peculiarities of airline networks, and a variety of solution algorithms have been proposed to accommodate the large scale of real-world airline operations. Starting from basic leg-based fleet assignment models (FAM), enhanced models have achieved a greater realism and wider acceptance among airlines by considering additional features of real-world airline networks (for a detailed overview, see Sherali et al. (2006)) such as itinerary-based demand (IFAM) (Barnhart et al. 2002), flexible departure time (Levin 1971, Rexing et al. 2000), and re-fleeting mechanisms (Sherali et al. 2005). More recently, researchers have focused on the development of integrated flight scheduling and fleet assignment models (Lohatepanont and Barnhart 2004, Barnhart et al. 2009, Sherali et al. 2010, 2013, Pita et al. 2013, 2014, Dong et al. 2016) that involve the simultane-

ous selection of flight legs to be included in the schedule and the allocation of aircraft types to these legs. The primary added value of using integrated models rather than following a sequential approach is that integrated modeling allows for the explicit consideration of supply-demand interactions, thus enabling superior decision making and improved schedule plans. Because scheduling and fleeting decisions take demand as a main input while demand is directly affected by the resulting network capacity (supply), separating market size estimation and seat capacity allocation fails to fully capture the reciprocal relationship between demand and supply, potentially resulting in sub-optimal decisions.

3.2.2   Supply-demand interactions

Generally speaking, there are two primary types of supply-demand interactions, both of which should be considered in airline planning. First, the overall level of air transport supply provided in an air transport market, i.e., an origin-destination (O&D) city pair, affects the total number of people willing to travel in that market (demand generation). In addition to the exogenous strength of the socio-economic attraction factors between two regions, the level of transportation provided plays a key enabling role in generating passenger traffic (Grosche et al. 2007, Adler et al. 2018, Boonekamp et al. 2018). Second, the specific service attributes of itineraries that an airline makes available in a given market affect the redistribution of passengers and eventually determine the portion of the market that the airline can capture from the competition (demand allocation) (Coldren et al. 2003).

Demand Allocation

Most previous contributions in this field have focused primarily on the demand allocation problem, as it is more sensitive to planning decisions. The typical approach has been to rely on a spill-and-recapture process (Lohatepanont and Barnhart 2004, Sherali et al. 2010, 2013) under which a fixed and exogenous demand is assumed for each itinerary and recapture rates, reflecting the likelihood that a passenger whose preferred itinerary is capacitated will pick an alternative one (see Kniker 1998), are

used to redirect "spilled" passenger flows. Advancements of this approach have involved the definition of unconstrained demand at the market level and the simultaneous estimation of an airline's market share as an "S-shaped" function of flight frequency (Vaze and Barnhart 2012, Pita et al. 2013).

Recently, researchers have been relying on discrete choice modeling to circumvent the primary limitations of the spill-and-recapture approach and to improve the degree of behavioral realism in the allocation of passengers over available itineraries. The theory of discrete choice modeling is rooted in random utility theory, in which passengers select their preferred itinerary in a given market according to the relative utilities provided by all available itineraries. In the air transport field, a number of studies have been conducted to investigate the factors that affect the itinerary choices of passengers, highlighting the important role played by connectivity-related attributes such as frequency, flight and connecting time, and the number of flight legs as factors that represent, along with airfares and departure times, the disutility of indirect vs. nonstop itineraries (see, e.g., Koppelman et al. 2008, Lurkin et al. 2017).

Although discrete choice models have often been used to support other stages of the planning process – e.g. revenue management (Talluri and van Ryzin 2004) –, the first contribution involving the application of discrete choice modeling to integrated flight scheduling and fleet assignment was Atasoy et al. (2014). In their model, scheduling, fleeting and pricing decisions are optimized altogether in a monopoly framework where passengers are distributed over alternative itineraries based on a multinomial logit (MNL) formulation. The result is a mixed-integer nonlinear model that significantly improves the treatment of supply-demand interactions, but is not applicable to realistic instances because of intractability.

Building on this limitation, Dong et al. (2016) developed a model in which the nonlinear MNL function is approximated by a set of linear constraints. This makes the model more tractable and solvable for larger instances and, although this linear representation is only equivalent to the MNL formulation in the absence of capacity constraints, it ensures that itinerary market shares are proportional to the attractiveness value and produces substitution patterns among itineraries that closely follow

the MNL formulation. Despite being very practical, this model suffers from the main limitation of MNL formulations, i.e., the Independence from Irrelevant Alternatives (IIA) property, which implies that the odds ratios between pairs of alternatives are independent of the presence of other alternatives and that the relative preferences between two alternatives are not affected by modifications of the choice set. The MNL formulation assumes independence between error terms and therefore fails to capture the correlation between alternatives, potentially yielding unrealistic substitution patterns.

From the literature on discrete choice, a typical approach to address the IIA property is to rely on nested logit formulations. Following this approach, Cadarso et al. (2017) developed a demand allocation model that accommodates the presence of intermodal competition and properly captures the patterns of correlation between and within airline types (FCC vs LCC). In a recent paper, Wei et al. (2020) referring to the optimization of integrated timetable and fleet assignment, presented an alternative approach to systematically tackling the IIA issue in the context of passenger itinerary choice based on a generalized attraction model (GAM) (Gallego et al. 2015).

Demand Generation

While the modeling of passenger distribution has been the subject of a number of studies, the systematic inclusion of trip generation dynamics into flight scheduling and fleet assignment models has been given much less attention. In most approaches, market demand is not allowed to change with the allocation of capacity and/or variations in the level of service in a specific market, thereby reinforcing the underlying assumption that leg selections do not affect market demand (Sherali et al. 2010).

Lohatepanont and Barnhart (2004) made the first contribution focusing on the importance of accounting for demand generation within flight scheduling and fleet assignment optimization. In their model, unconstrained demand for each market is computed using a schedule evaluation model (SEM) assuming that all itineraries are flown. Demand correction terms are then introduced to account for network effects in which demand spill is associated with the deletion of each optional itinerary from the network configuration. Although correction terms could theoretically be estimated by

running the evaluation package as many times as the number of possible combinations of flight additions and deletions, this approach is practically unfeasible, and a solution heuristic was proposed to handle the model.

In the model developed by Pita et al. (2014), demand per market is linearly adjusted through elasticities to account for changes in the average airfare. This approach directly encompasses the demand-stimulating effect of lower prices but fails to account for other factors such as increased frequency and lower travel time and its effectiveness is conditioned on the ability of the target airline to influence the average market price.

Models relying on discrete choice for passenger allocation can include a "no-fly option" to account for the fact that not all potential demand is currently served (Dong et al. 2016). In an MNL framework, however, choosing not to fly is considered to be equivalent to choosing among alternative itineraries and, consequently, the problem of IIA is exacerbated. To clarify, let us consider an aviation version of the blue-bus/red-bus paradox, in which there are two itineraries—A and B (which are nearly identical) —and the current market is split as follows: 20% itinerary A, 20% itinerary B, and 60% no-fly option. If itinerary A is dropped from the choice set, according to the MNL formulation 5% of passengers would re-route to itinerary B while 15% would stop flying, maintaining a constant odds ratio between B and the no-fly option. However, given the similarity between itineraries A and B, it would be more reasonable to expect that most of itinerary A's passengers would opt for itinerary B rather than stop flying.

An improved approach to addressing the trip generation problem was presented in Cadarso et al. (2017), who estimated the overall market demand using a multiplicative model as a function of the average transport price, travel time, and total frequency. As these attributes are the same decision variables used by the model that determines airline choices in a given market (allocation model), the two steps were combined and solved simultaneously. This model was applied in the context of frequency planning and approximate timetable development on a pure hub-and-spoke network and operates under the assumptions that a unique route is provided by the

airline in each city-pair market such that average attributes (price and total travel time) can be considered for each time period.

3.2.3   Paper contributions

The aim of this paper is to develop an integrated optimization framework that consistently accounts for supply-demand interactions—both demand generation and allocation—within itinerary-based integrated flight scheduling and fleet assignment models. Whereas previous works focused primarily on demand allocation and considered aggregate models to estimate variations in total market demand, in our approach the utility is defined at the itinerary level to simultaneously model competition among air travel itineraries and improve the assessment of air trip generation. This allows us to flexibly consider situations in which an airline provides different routes in a certain market (e.g., multi-hub systems, hub bypass flights) and appraise the contribution of specific itinerary attributes to demand generation. In detail, this paper makes the following contributions. First, we develop a mixed-integer programming model for flight scheduling and fleet assignment, which incorporates a utility-based formulation for air travel demand in a novel way. Specifically, we leverage an aggregate nested logit formulation to enable a better and more realistic treatment of supply and demand interactions by considering both demand allocation and generation dynamics simultaneously. Second, we propose and validate a tailored piecewise linearization scheme based on least squared fit to effectively approximate the nonlinear demand functions. Third, we investigate the applicability of the proposed modeling and algorithmic framework by means of computational experiments involving randomly-generated realistic problem instances of different sizes. The results reveal that the model can handle mid-size networks and provide useful insights for larger cases in reasonable time. Fourth, we illustrate the benefits and practical insights that can be derived from the application of the proposed approach for the domestic and international networks of Alitalia (AZ) based on Milan-Linate (LIN) airport. We show how the model can deliver significant profit increases compared to the baseline scenario and better unveil opportunities for demand stimulation compared to an inelastic market

demand scenario that replicates the use of basic MNL models.

3.3   Model Formulation

In this section, we present the formulation of the proposed *Integrated demand Generation and Allocation flight Scheduling and Fleet Assignment Model (IGASFAM)*. In Section 3.3.1, we describe the nested logit formulation used to represent itinerary and total market demand. In Section 3.3.2, we present the notation and formulation of the optimization model, along with the underlying assumptions. In Section 3.3.3, we illustrate the the piecewise linearization scheme implemented to address the nonlinearities in the model.

3.3.1   Demand model

The demand model is formulated as an aggregate nested logit model inspired by the works of Wei and Hansen (2005) and Hsiao and Hansen (2011). The nesting structure is based on the separate grouping of air travel itineraries and the non-air alternative, i.e., the "outside good," to provide a choice condition that comprehends both the no-travel option and traveling via other modes (Berry and Jia 2010). This pursues the twofold aim of overcoming the primary shortcomings of the IIA property[1] in the context of air transport and making demand generation elastic by dynamically incorporating the composite utility provided by air travel itineraries into the trip generation stage.

For each market, the saturated demand, given as the theoretical maximum number of trips that can be expected in a given market (city-pair) if there were no travel impedance, is first broken down into the non-air alternative and the air travel nest, which is further decomposed into the different air travel itineraries available to passengers. As such, this nested logit formulation simultaneously describes competition across itineraries (demand allocation) and their composite contribution to air travel

---

[1]Here, a general formulation with only two nesting levels is presented. However, itineraries can be further grouped to allow for more flexible substitution patterns, for instance, between single airlines, alliances, or legacy and low-cost carriers.

utility, which in turn competes with the non-air alternative to capture a higher portion of demand (demand generation). Following the customary procedure (Berry 1994), the utility of the outside good is normalized to zero, $V_{non-air} = 0$, while the composite utility of the air travel nest can be generally stated as the sum of a market-specific component (δ) and the compound utility from lower nests, here represented by the inclusive value (IV), i.e., $V_{air} = \delta + \theta IV$, where $IV = \log \sum_{j \in I} e^{V_j/\theta}$, I is the set of itineraries available, and V_j is the utility of itinerary j. The nesting coefficient (θ) represents the degree of inter- and intra-correlation between and within nests, respectively. In practical terms, θ governs substitution of air travel for the outside good, thereby quantifying the extent to which changes in air transport supply affect total air travel demand. For a given market, let Q be the total market demand for air travel and q_i be the demand for itinerary i. Consistent with the nested logit formulation, Q and q_i can be synthetically expressed as follows:

$$Q = \frac{D}{1 + \gamma e V_{tot}^{-\theta}} \tag{3.1}$$

$$q = \frac{D}{\gamma e V_{tot}^{1-\theta} + e V_{tot}} e V_i \tag{3.2}$$

where D is the market saturated demand, $eV_{tot} = \sum_{j \in I} e^{V_j/\theta}$, $eV_i = e^{V_i/\theta}$, and $\gamma = 1/e^{\delta}$. In Section 3.5.1, we empirically estimate D as the geometric mean between the populations at the market endpoints multiplied by a proportionality factor and consider a linear-in-parameters itinerary utility depending on key supply-side attributes, including flight and connecting time, type of service (nonstop vs connecting itinerary), and average price. Functions describing Q and q_i have desirable characteristics. As θ is bounded between zero and one for consistency with the utility-maximizing theory and γ is strictly positive, Q increases with eV_{tot} and is concave downward, while q_i is a monotonic decreasing function of eV_{tot} (Figure 3-1). It follows that the demand stimulation effect resulting from improving air transport (by, for instance, lowering prices or increasing capacity) will be lower if a market is already well-served and less than proportional so that itinerary demand still decreases but less

sharply than if induced demand is ignored. Through the estimation of market-specific parameters, different degrees of trip generation elastic response can be accommodated to account for the specificities of each market and implicitly consider the intensity of intermodal competition[2].
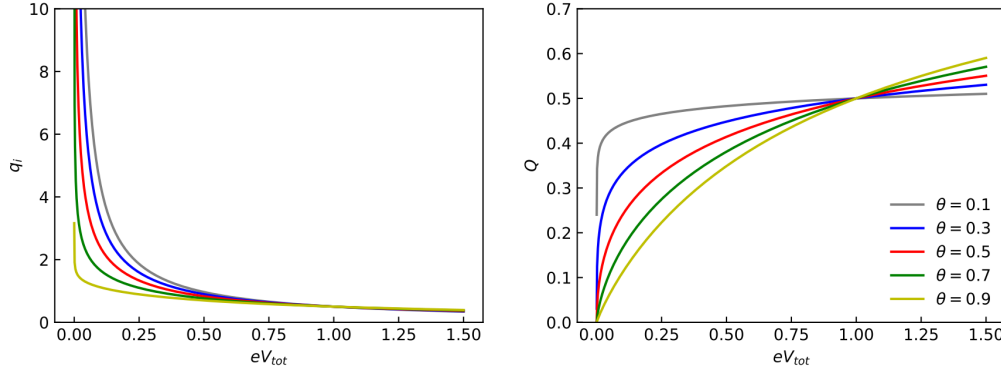
Figure 3-1: Itinerary demand function q_i (left) and total market demand Q (right) with $D = 1$, $eV_i = 1$, and $\gamma = 1$.

3.3.2 Optimization model

The optimization model is formulated as a multi-commodity network flow model in which commodities are represented as fleet types to be allocated on a time-space network (Figure 3-2). Similar to Dong et al. (2016), the time-space network is composed of activity nodes, which represent events, i.e., flight arrivals or departures at a certain airport, connected by arcs. Arcs can be of two types: flight arcs representing potential flights that connect activity nodes at different airports; and ground arcs, which link nodes at the same location, representing aircraft stopped at a specific airport. Feasible travel alternatives for passengers are represented by itineraries, which either coincide with single flight arcs (nonstop itineraries) or comprise a combination of flight arcs (connecting itineraries).

---

[2]While it would be desirable to have explicit data on alternative transport modes for each market, these data are typically difficult to gather systematically and airlines themselves often lack this information on a full network scale. Hence, despite the inherent limitation of assessing demand variations from changes in specific attributes of competing modes, this modeling approach has the practical advantage of consistently reflecting the as-is intermodal competition through market specific demand curves.
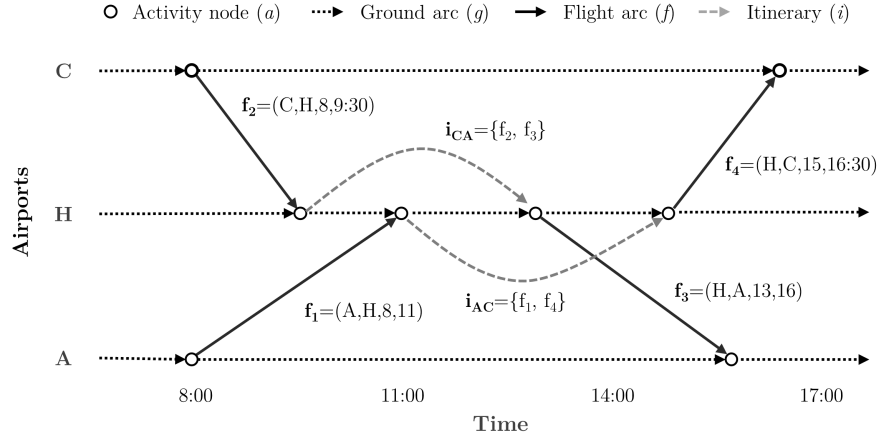
Figure 3-2: Time-space network representation of a simple A-H-B hub and spoke network.

Before introducing our model notation, we clarify the main assumptions on which the model is based:

- *Competition*: Demand is computed according to the aggregate nested logit model outlined in Section 3.3.1. As such, static competition is directly considered in the estimation of demand, while we ignore the strategic reaction of competing airlines. We assume that the characteristics of the competition, such as the schedule and average price, are known and fixed. This is in line with the literature on flight scheduling and fleet assignment due to the focus on a single airline and large-scale networks that encompass several markets. To address this issue, an iterative approach to investigate the competitive dynamics in a game-theoretical fashion could be implemented (see, e.g., Adler et al. 2010, Cadarso et al. 2017), but this is out of the scope of this paper.

- *Pricing*: For each itinerary, we assume that average prices for different passenger types are given. Some researchers have tried to endogenize pricing decisions into flight scheduling and fleet assignment models (see, e.g., Atasoy et al. 2014, Dong et al. 2016). However, this leads to severe tractability issues. In addition, given the tactical nature of IGASFAM, taking itinerary prices as exogenous inputs is reasonable and widely accepted in the literature, as it indeed reflects current

practice.

- *Time-of-day preferences*: In our model, the consideration of demand patterns throughout the day is not achieved by dividing the planning horizon into demand periods (as is done, for example, by Barnhart et al. (2002), and Pita et al. (2013, 2014)). As formulated, the IGASFAM can accommodate different demand distributions by augmenting the utility function to include terms that capture time-of-day preferences, by using, for example, temporal dummies (Coldren et al. 2003), continuous departure time of day preference curves (Lurkin et al. 2017), or schedule delay functions (Koppelman et al. 2008).

- *Operational considerations*: We consider one single day of operations and assume repeatability. Continuity in the network is ensured by introducing a "wrap-around arc" that connects the last and first events at each airport. Using a single day of operations is in line with the empirical setting (see Section 3.5) and the definition of utilities at the itinerary level, but extending the model to other time horizons would be rather straightforward. Slot and other airport-specific constraints are explicitly modeled by defining a set of airport-period of time combinations characterized by a maximum number of aircraft movements. Eventually, turnaround time and other operational considerations are directly accounted for in the structuring of the multicommodity time-space network.

Next, we introduce the model notation and formulate the IGASFAM mathematically.

Decision Variables

$x_{af} \in \{0,1\}$	=1 if aircraft type a is assigned to flight arc f, = 0 otherwise
$y_{ag} \in \mathbb{N}^+$	the flow value of aircraft type a through ground arc g
k_i	itinerary status variable (1 if itinerary i is operated, 0 otherwise)
eV_{mh}^{tot}	total exp-utility in market m for passenger type h
d_{mh}	equivalent demand in market m for passenger type h
φ_{ih}	the flow of passengers of type h accommodated on itinerary i

Inputs: sets and parameters

T	set of event times indexed by t
AT	set of fleet types indexed by a
N	set of activity nodes indexed by n
M	set of markets (city-pairs) indexed by m
F	set of flight arcs indexed by f
G	set of ground arcs (including wrap-around arcs) indexed by g
I	set of itineraries indexed by i
H	set of passenger types indexed by h
S	set of slots indexed by s
F^m	set of mandatory flight arcs indexed by f
N_o	set of activity nodes belonging to airport o
CL_f	set of flight arcs crossing the count line indexed by f
CL_g	set of ground arcs crossing the count line indexed by g
F_n^{in}	set of inbound flight arcs to activity node n indexed by f
F_n^{out}	set of outbound flight arcs from activity node n indexed by f
G_n^{in}	set of inbound ground arcs to activity node n indexed by g
G_n^{out}	set of outbound ground arcs from activity node n indexed by g
I_m	set of itineraries in market m indexed by i
F_i	set of flight arcs belonging to itinerary i indexed by f
I_f	set of itineraries requiring flight arc f indexed by i
F_s	set of flight arcs using slot s
$fleet_a$	fleet size by aircraft type
cap_a	adjusted seat capacity by aircraft type accounting for a maximum allowable load factor
$nlegs_i$	number of flight legs in itinerary i
eV_{ih}	exp-utility of itinerary i for passenger type h
eV_{mh}^{fix}	exogeneous exp-utility provided by competing airlines in market m for passenger type h
D_{mh}	saturated demand in market m for passenger type h
θ_{mh}	nesting coefficient in market m for passenger type h
γ_{mh}	market-specific constant for market m and passenger type h
p_{ih}	average ticket price on itinerary i for passenger type h
c_{af}^{op}	operating costs for an aircraft of type a serving flight arc f
c_f^{pax}	per passenger airport costs for flight arc f
c_{af}^{mov}	per movement airport costs for flight arc f
mov_s	slot capacity, defined as the maximum number of movements that can be accommodated in slot s

Mathematical formulation

$$\max \quad \sum_{i \in I} \sum_{h \in H} \varphi_{ih} p_{ih} - \sum_{f \in F} \sum_{i \in I_f} \sum_{h \in H} c_f^{pax} \varphi_{ih} - \sum_{a \in AT} \sum_{f \in F} (c_{af}^{op} + c_{af}^{mov}) x_{af} \tag{3.3}$$

$$\sum_{f \in F_n^{in}} x_{af} + \sum_{g \in G_n^{in}} y_{ag} = \sum_{f \in F_n^{out}} x_{af} + \sum_{g \in G_n^{out}} y_{ag} \qquad \forall n \in N, \forall a \in AT \tag{3.4}$$

$$\sum_{f \in CL_f} x_{af} + \sum_{g \in CL_g} y_{ag} \leq fleet_a \qquad \forall a \in AT \tag{3.5}$$

$$\sum_{a \in AT} x_{af} = 1 \qquad \forall f \in F^m \tag{3.6}$$

$$\sum_{a \in AT} x_{af} \leq 1 \qquad \forall f \in F^o \tag{3.7}$$

$$k_i - \sum_{a \in AT} x_{af} \leq 0 \qquad \forall f \in F_i, \forall i \in I \tag{3.8}$$

$$k_i - \sum_{a \in AT} \sum_{f \in F_i} x_{af} \geq 1 - nlegs_i \qquad \forall i \in I \tag{3.9}$$

$$eV_{mh}^{tot} = eV_{mh}^{fix} + \sum_{i \in I_m} eV_{ih} k_i \qquad \forall m \in M, \forall h \in H \tag{3.10}$$

$$d_{mh} = D_{mh} / [\gamma_{mh} (eV_{mh}^{tot})^{1-\theta_{mh}} + eV_{mh}^{tot}] \qquad \forall m \in M, \forall h \in H \tag{3.11}$$

$$\varphi_{ih} \leq d_{mh} eV_{ih} \qquad \forall i \in I, \forall h \in H \tag{3.12}$$

$$\sum_{h \in H} \sum_{i \in I_f} \varphi_{ih} \leq \sum_{a \in AT} x_{af} cap_a \qquad \forall f \in F \tag{3.13}$$

$$\sum_{f \in F} \sum_{a \in AT} x_{af} \leq mov_s \qquad \forall s \in S \tag{3.14}$$

$$x_{af} \in \{0, 1\} \qquad \forall a \in AT, \forall f \in F \tag{3.15}$$

$$y_{ag} \in \mathbb{N}^+ \qquad \forall a \in AT, \forall g \in G \tag{3.16}$$

$$k_i \in \mathbb{R}^+ \qquad \forall i \in I \tag{3.17}$$

$$eV_{mh}^{tot} \in \mathbb{R}^+, d_{mh} \in \mathbb{R}^+ \qquad \forall m \in M, \forall h \in H \tag{3.18}$$

$$\varphi_{ih} \in \mathbb{R}^+, \qquad \forall i \in I, \forall h \in H \tag{3.19}$$

The objective function (3.3) maximizes the expected operating profits, i.e., the difference between airline revenues and costs. The revenue term is the product of the average fare per itinerary and the number of passengers accommodated in each itinerary. Costs are split into three main categories: variable flight costs proportional to the number of passengers (airport passenger fees), fixed flight costs (per-movement airport fees), and direct operating flight costs dependent on the aircraft type being used. Constraints (3.4)–(3.5) are basic network constraints that ensure flow balance and resource availability, respectively. Constraints (3.4) force the number of departures and arrivals per aircraft type at any activity node to be equal, while constraints (3.5) count the aircraft flows crossing the count line and force this value to be lower than the actual per-type aircraft availability. Given the presence of wrapped-around arcs, Constraints (3.4)–(3.5) suffice to ensure schedule repeatability within the planning period. Constraints (3.6)–(3.7) are coverage constraints ensuring that flight arcs are flown by at most one aircraft; specifically, mandatory flights must be flown (3.6), while optional ones may or may not be flown (3.7). Constraints (3.8)–(3.9) are itinerary status constraints under which the itinerary variable k_i is set equal to one when itinerary i is operated (3.9), i.e., if all its flight legs are flown; alternatively, if at least one flight leg is not flown, constraints ((3.8) force k_i to be equal to zero. The next constraints (3.10) compute the total exponential utility for each market eV_{mh}^{tot}, given by the sum of the exogenous utility eV_{mh}^{fix}, i.e., the utility provided by competing airlines, and the endogenous utility, given by the sum of the exp-utilities of activated itineraries. Constraints (3.11) takes eV_{mh}^{tot} and compute what we call equivalent demand (d_{mh}), that is, the estimated demand for an itinerary for which exp-utility equals one in market m. Constraints (3.11) are nonlinear but can be properly approximated by piecewise linearization (Figure 3-3a), as explained below. Constraints (3.12)–(3.13) ensure that the accepted demand on each itinerary does not exceed the demand estimated by the model (Figure 3-3b) and enforce a capacity constraint under which the number of total passengers on each flight arc is lower than the allocated adjusted seat capacity (cap_a) defined as the aircraft nominal seat capacity multiplied by a maximum allowable load factor to reflect demand uncertainty and the function-

ing of revenue management systems. Constraints (3.14) enforce that the number of flights using a given slot does not exceed its capacity. Finally, Equations (3.15)–(3.19) define the domain of the variables.

Once the model is solved, second-order passenger spill and recapture effects can be considered by iteratively redistributing spilled demand over activated itineraries according to the integrated demand model and dropping capacitated itineraries at each step, until the unaccommodated demand is zero or (adjusted) capacity is saturated. This approach follows the intuition that market demand is eventually and simultaneously determined by the level of allocated supply. Hence, for a given supply configuration, i.e., a given set of itineraries with their own utility values, recapture rates on non-activated itineraries should be zero while recapturing on activated itineraries can be determined using the integrated demand model to consistently represent passengers that will stop traveling because the preferred alternative is no longer available or will instead re-route to other itineraries. This approach is justified by two main considerations: (1) The demand model underpinning the IGASFAM implicitly considers capacity; although Q theoretically represents the unconstrained, or true, demand for air transportation in a given market, it is empirically estimated based on observable demand (i.e., accommodated passengers including those spilled and recaptured), providing an estimate of realized demand that partly reflects capacity constraints and commonly used aircraft. For this reason, recent papers using discrete choice modeling to represent passenger demand do not consider these second-order spill and recapture effects (e.g., Dong et al. 2016, Wei et al. 2020); (2) The utility of the no-fly alternative is likely to be large compared to air travel (Wei and Hansen 2005). From these two considerations, it follows that recaptured demand will unlikely change the optimal solution in terms of activated itineraries. Sensitivity analyses were performed to ensure that the explicit consideration of spill and recapture in each market separately never leads to a different optimal solution. We can thus consider spill and recapturing once the optimization is solved to marginally adjust the passenger figures on activated itineraries.

3.3.3 Solution methods

Piecewise linearization

Constraints 3.11 are not linear, thus making the IGASFAM a mixed integer nonlinear problem (MINLP). To solve it, we linearize constraints 3.11 using market-specific piecewise linear functions. As the IGASFAM assumes different saturation curves for each market and continuously defined utility functions, using the same piecewise linearization scheme for each O-D pair is not recommended. To improve the accuracy of the approximation and reduce the computational burden, a domain reduction strategy is adopted and a preprocessing step is performed to optimally define the location of breakpoints.

For each market and passenger type (subscripts m and h omitted), the exogenous utility eV_{fix} and the set of optional itineraries I are known, making it possible to compute the maximum exponential utility allocated to the market by assuming that all itineraries are operated, $eV_{max} = eV_{fix} + \sum_{i \in I} eV_i$. Therefore, the demand curve is not to be approximated in its entire domain but can be bounded between these two values without eliminating any feasible solution (Figure 3-3a). Furthermore, we can map all possible combinations of itinerary additions by considering the set of all subsets of I (referred to as $\mathcal{P}$, indexed by p) and compute the corresponding total utility $(eV_p^{tot} = eV_{fix} + \sum_{i \in I_p} eV_i)$ and equivalent demand value (d_p). Finally, given the feasible region and the number of linear pieces, the breakpoint locations are selected to maximize the fit across these points $\{(eV_p^{tot}, d_p) \ \forall p \in \mathcal{P}\}$ and thus minimize the approximation errors.

We now describe the procedure utilized to determine the optimal location of breakpoints. Let η be the number of breakpoints, which partition the bounded domain into $\eta - 1$ segments. For each market, we solve the following bilevel optimization problem:

$$\underset{\mathbf{b} \in \mathcal{B}}{\arg \min} \, RSS(\mathbf{b}) \tag{3.20}$$

$$RSS(\mathbf{b}) = \min_{\mathbf{y} \in \mathbb{R}^\eta} \sum_{j \in \mathcal{L}_{\mathbf{b}}} \sum_{p \in \mathcal{P}_j} (\hat{d}_j(eV_p^{tot}, y_j, y_{j+1}) - d_p)^2 \tag{3.21}$$
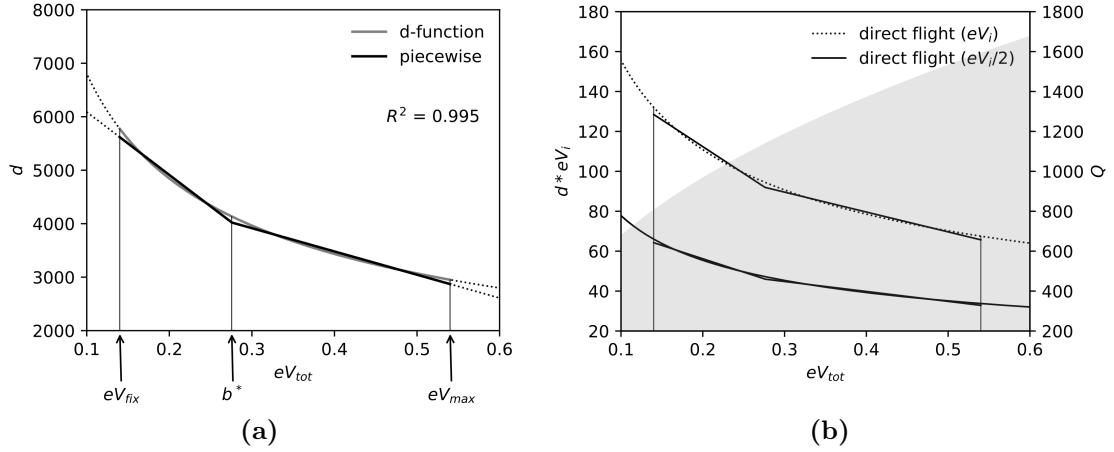
Figure 3-3: Example of piecewise linearization for the Milan-Rome market in May 2018: (a) Piecewise linearization of constraint (3.11); (b) Itinerary demand for an average direct itinerary with exp-utility eV_i (dotted line) and $eV_i/2$ (dashed line), respectively, and total market demand (shaded area) as a function of total exp-utility in the market eV_{tot}.

The outer optimization model (Equation 3.20) determines the optimal set of breakpoints (on the x-axis), $\mathbf{b} = (b_1, ..., b_\eta) \in \mathcal{B}$, where $\mathcal{B} = \{(b_1, ..., b_\eta) \in \mathbb{R}^\eta : b_1 = eV_{fix}, b_\eta = eV_{max}, eV_{fix} \leq b_k \leq eV_{max} \ \forall \ k = 1, ..., \eta\}$, that minimize the overall sum-of-squares of the residuals ($RSS(\mathbf{b})$). To solve this model we use the Differential Evolution (DE) algorithm proposed by Storn and Price (1997) which proceeds iteratively by updating solution candidates that improve the objective function. At each iteration, the evaluation of the objective function constitutes the inner optimization model (Equation 3.21), wherein a least squares fit is performed to determine the breakpoint values (on the y-axis), $\mathbf{y} = (y_1, ..., y_\eta)$, of the best continuous piecewise linear model for a given set of breakpoint locations (Kundu and Ubhaya 2001, Jekel and Venter 2019). In Equation 3.21, $\mathcal{L}_\mathbf{b}$ represents the set of intervals identified by the set of breakpoints $\mathbf{b}$, $\mathcal{P}_j$ is the subset of itinerary subsamples that belong to interval $j \in \mathcal{L}_\mathbf{b}$ and $\hat{d}_j(eV_p^{tot}, d_j, d_{j+1})$ are the fitted values in interval j.

We set the number of linear segments equal to two as this allows us to achieve a sufficient level of accuracy (above 94%, refer to Section 3.5.4 for details) and apply a convex combination formulation. Let $\mathbf{b}_{mh}$ be the set of optimal breakpoints for market m and passenger type h, obtained by solving the optimization model outlined

105

in Equations (3.20)-(3.21). We define $\mathcal{L}_{\mathbf{b}_{mh}}$ as the set of corresponding intervals and ψ_{bj} as a parameter equal to 1 if breakpoint b belongs to interval j, and 0 otherwise. We further introduce λ_b as a non-negative variable defined over the set of breakpoints, representing the weights in the convex combination. Eventually, τ_j is a binary variable equal to 1 if interval j is activated, and 0 otherwise. Using this notation, we replace the nonlinear constraints (3.11) with the following linear constraints (see Vielma et al. (2010) for more details).

$$eV_{mh}^{tot} = \sum_{b \in \mathbf{b}_{mh}} \lambda_b eV_b^{tot} \qquad \forall m \in M, \forall h \in H \tag{3.22}$$

$$d_{mh} = \sum_{b \in \mathbf{b}_{mh}} \lambda_b \{D_{mh}/[\gamma_{mh}(eV_b^{tot})^{1-\theta_{mh}} + eV_b^{tot}]\} \qquad \forall m \in M, \forall h \in H \tag{3.23}$$

$$\sum_{b \in \mathbf{b}_{mh}} \lambda_b = 1 \qquad \forall m \in M, \forall h \in H \tag{3.24}$$

$$\sum_{j \in \mathcal{L}_{\mathbf{b}_{mh}}} \tau_j = 1 \qquad \forall m \in M, \forall h \in H \tag{3.25}$$

$$\lambda_b \leq \sum_{j \in \mathcal{L}_{\mathbf{b}_{mh}}} \tau_j \psi_{jb} \qquad \forall b \in \mathbf{b}_{mh}, \forall m \in M, \forall h \in H \tag{3.26}$$

$$\lambda_b \geq 0 \qquad \forall b \in \mathbf{b}_{mh}, \forall m \in M, \forall h \in H \tag{3.27}$$

$$\tau_j \in \{0,1\} \qquad \forall j \in \mathcal{L}_{\mathbf{b}_{mh}}, \forall m \in M, \forall h \in H \tag{3.28}$$

Tightening constraints

Following the piecewise linearization of constraints 3.11, the IGASFAM reduces to a large mixed-integer problem. To improve its solvability, we consider a set of valid inequalities to tighten the model representation and accelerate convergence to optimality. Specifically, we add the set of tightening constraints outlined in Proposition 6, which bound the itinerary demand to be less than or equal to its maximum potential value (i.e., if the itinerary is the only one added to the market) (Eq. 3.30) and the maximum aircraft capacity.

Proposition 6 *The following is a valid inequality for the IGASFAM.*

$$\varphi_{ih} \leq \min(\tilde{\varphi}_{ih}, \max_{a \in AT} cap_a) k_i \qquad \forall i \in I, \forall h \in H \qquad (3.29)$$

where

$$\tilde{\varphi}_{ih} = \{D_{mh}/[\gamma_{mh}(eV_{mh}^{fix} + eV_{ih})^{1-\theta_{mh}} + (eV_{mh}^{fix} + eV_{ih})]\}eV_{ih} \qquad (3.30)$$

Proof of Proposition 6

The validity of constraint (3.29) directly follows from previous constraints based on an approach similar to that used in Sherali et al. (2010). If $k_i = 0$, than $\exists f \in F_i | \sum_{a \in AT} x_{af} = 0$ (3.8)–(3.9), therefore (3.29) is valid by (3.13). If $k_i = 1$, then $eV_{mh}^{tot} \geq eV_{mh}^{fix} + eV_{ih}$ (3.10) and $\varphi_{ih} \leq d_{mh}eV_{ih} \leq \tilde{\varphi}_{ih}$ (3.11), such that $\varphi_{ih} \leq \tilde{\varphi}_{ih}$. In addition, (3.13), along with (3.8) and (3.9), imply that $\varphi_{ih} \leq \max_{a \in AT} cap_a$. This completes the proof. ∎

In addition to the valid cuts given in Proposition 6, we tested other cuts (e.g., Sherali et al. 2010, 2013) but without any further significant computational benefits.

3.4 Computational Experiments

In this section, the computational tractability of the proposed model is examined based on the results of numerical experiments using randomly generated problem instances of various sizes. The objective is to demonstrate the applicability of the model on real-world-sized networks and to test its computational performance under different sensible assumptions. We consider a pure hub-and-spoke network and test different network sizes (500, 1,000, and 2,000 city-pair markets), fleet compositions (one or two fleet types), and network constraints in terms of percentage of mandatory flights (0 or 50%). For each network size, five prototypal networks were generated to produce 60 (3 × 2 × 2) model instances in total.

To ensure the meaningfulness of the generated scenarios, market characteristics, such as market distance and endpoint populations, competing services, and costs were drawn from empirical distributions on an intra-European basis[3]. For each point-to-point market, we considered eight to ten optional flight legs randomly allocated throughout the day. One-stop itineraries were built considering the shortest path and applying a threshold of below 1.5 on the routing factor (Paleari et al. 2010)[4]. Utilities were computed according to the model outlined in Section 3.5.1, while average prices were computed using the fare model proposed by Lieshout et al. (2016) and randomized following a normal distribution. All numerical experiments were carried out using the CPLEX 12.9 MILP solver on an Intel(R) Core(TM) i7-8700K CPU with a frequency of 3.70 GHz and 32 GB of RAM, and the maximum computational time was set to two hours. The piecewise least square fit method was implemented using the pwlf python package and entailed short preprocessing times (less than 5 minutes) across all instances.

---

[3]Refer to Section 3.5 for details on data available.

[4]Inconvenient routing substantially undermines an airline's ability to serve a market through connecting flights (Burghouwt and Redondi 2013). As a result, the number of relevant city-pair markets tends to be lower than its maximum value. For instance, if a threshold of below 1.5 is applied to the routing factor, a hub-and-spoke network with 50 airports—assuming, for simplicity, independent catchment areas— will correspond to an average of 500 markets, which is far below the potential number of 2,450 (50 $times$ 49). This directly follows from the hub location and geographic dispersion of demand, which is in line with real-world network configurations, as depicted in Figure 3-4.

Table 3.1 summarizes the results obtained. For each class of instances, Table 3.1 reports summary statistics on the size of the instances, including the number of markets (# *mkts*), the number of variables (# *cols*) and the number of constraints (# *rows*), along with the computational time (*Time*) in seconds and optimality gaps (*Gap*). In general, it was possible to solve to optimality, or quasi-optimality, all of the instances considered except the most difficult ones—involving the largest network size (100 airports), two fleet types, and no mandatory flights. For these instances, the optimality gaps could be significantly reduced by increasing the computation time to five hours—from 14.4% to 5.9%, on average (values into brackets). Note that, as we are dealing with tactical planning, time is not a scarce resource and five hours is still an acceptable time threshold to compute quality solutions.

The experimental results reveal that the model can handle pure scheduling problems fairly well, even if starting from a blank schedule. As the number of mandatory flights increases, the number of combinations from which the most profitable mix of optional itineraries can be selected decreases, thereby reducing the computational effort required to solve the model. This effect is exacerbated by the similarity of utility values among optional itineraries, which tends to generate a larger set of similarly good solutions. Having different fleet types increases the computational burden because it essentially leads to interacting parallel time-space networks—one for each aircraft type—that enlarge the solution space of the model.

To evaluate whether the cases considered above properly reflect the real relevant network size of a company, intended as the number of markets that are strategically managed by an airline in their planning and not simply formed by sporadic connections, we consider three well-known European airlines: TAP (TP), Alitalia (AZ), and Lufthansa (LH). Their continental networks handled 8.8, 13.7, and 29.7 million passengers in 2018, respectively. Figure 3-4 shows the relation between the number of passengers and the number of markets. For TAP and Alitalia, the first 500 largest markets represent approximately 96.5 and 91.7% of their passengers, respectively, while the first 1,000 represent about 99.2 and 96.7%, respectively. The remaining markets are numerous and very thin, with a yearly volume of passengers below 1,000

(i.e., less than three passengers per day). In the case of Lufthansa, the number of relevant markets is much larger, with the first 2,000 accounting for 81.2% of total passengers. Therefore, the results of our computational experiments suggest that the model can handle midsize networks, such as those of TAP and Alitalia, within a reasonable time, and provide useful insights for networks as large as that of Lufthansa.
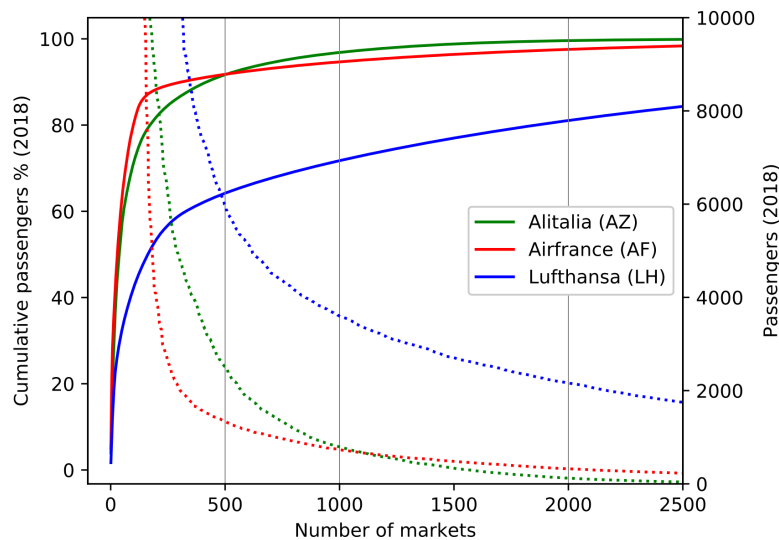
Figure 3-4: Real-world airlines' continental networks.

Markets are sorted from largest to smallest according to the volume of passengers. Solid lines represent the cumulative distribution of passengers (left-hand y-axis), while dotted lines represent the number of yearly passengers per market (right-hand y-axis).

3.5 Real-world study

In this section, we describe a study carried out to demonstrate the benefits that can be derived from the application of the proposed model to the Alitalia (AZ) continental flight network based at the Milan-Linate (LIN) airport. Alitalia is the main Italian flag carrier, which has been under special administration since May 2017. In 2018, Alitalia carried 21.5 million passengers. Its network has a main hub in Rome Fiumicino (FCO) and a base at Milan Linate (LIN). LIN serves one of the most populated areas

Table 3.1: Computational experiments

# Airports	Aircraft Types	Mandatory flights	Field	# mkts	# cols	# rows	Time (s)	Gap (%)
50	1	0%	*min*	462	11178	20710	168	0.00%
50	1	0%	*mean*	492	11785	21956	584	0.00%
50	1	0%	*max*	514	12269	22948	1225	0.00%
50	1	50%	*min*	462	11178	20710	13	0.00%
50	1	50%	*mean*	492	11785	21956	33	0.00%
50	1	50%	*max*	514	12269	22948	54	0.00%
50	2	0%	*min*	462	13075	21831	TL	1.0% (0.8%)
50	2	0%	*mean*	492	13676	23071	TL	1.9% (1.4%)
50	2	0%	*max*	514	14171	24064	TL	3.4% (2.3%)
50	2	50%	*min*	462	13075	21831	398	0.00%
50	2	50%	*mean*	492	13676	23071	818	0.00%
50	2	50%	*max*	514	14171	24064	2227	0.00%
70	1	0%	*min*	882	20153	38043	1077	0.00%
70	1	0%	*mean*	1017	22803	43424	4152	0.40%
70	1	0%	*max*	1124	25062	48075	TL	0.90%
70	1	50%	*min*	882	20153	38043	34	0.00%
70	1	50%	*mean*	1017	22803	43424	114	0.00%
70	1	50%	*max*	1124	25062	48075	181	0.00%
70	2	0%	*min*	882	22793	39592	TL	1.1% (0.6%)
70	2	0%	*mean*	1017	25419	44947	TL	3.5% (2.6%)
70	2	0%	*max*	1124	27663	49581	TL	7.2% (5.8%)
70	2	50%	*min*	882	22793	39592	3375	0.00%
70	2	50%	*mean*	1017	25419	44947	6293	0.50%
70	2	50%	*max*	1124	27663	49581	TL	1.50%
100	1	0%	*min*	1764	38580	74131	2972	0.00%
100	1	0%	*mean*	1928	41663	80313	6359	1.30%
100	1	0%	*max*	2170	46421	89935	TL	3.60%
100	1	50%	*min*	1764	38580	74131	101	0.00%
100	1	50%	*mean*	1928	41663	80313	1693	0.00%
100	1	50%	*max*	2170	46421	89935	4535	0.00%
100	2	0%	*min*	1764	42287	76275	TL	2.7% (0.9%)
100	2	0%	*mean*	1928	45352	82435	TL	14.4% (5.9%)
100	2	0%	*max*	2170	50089	92038	TL	37.4% (11.7%)
100	2	50%	*min*	1764	42287	76275	TL	0.00%
100	2	50%	*mean*	1928	45352	82435	TL	1.30%
100	2	50%	*max*	2170	50089	92038	TL	4.00%

Values into brackets refer to optimality gaps obtained with a computation time of 5 hours.
"TL" stands for time-limit.

in Europe[5]. This airport is part of a multi-airport system that also includes Milan Malpensa (MXP) and Milan Bergamo (BGY). The LIN catchment area significantly overlaps those of MXP and BGY (by 70 and 63%, respectively), leading to strong

---

[5]Stretching over central Lombardy and reaching up to eight million inhabitants within a one-hour driving time (best-guess statistics from Google Maps following Birolini et al. (2019)).

competitive interactions. Among the three airports, LIN is the closest to the Milan city center (7 km) and, since 2001, LIN has been subject to slot constraints and traffic distribution rules to limit its environmental burden and foster the development of intercontinental services at Milan Malpensa (MXP) [6]. Alitalia historically owns the majority of LIN's slots, which, given the richness of the catchment area and scarcity of resources, represent a valuable strategic asset. Alitalia's priority concern is the planning and design of a profitable network while retaining the slots at LIN (new flights cannot be added but only swapped using currently available slots), taking into account the competition from other carriers, both LCCs and FSCs, on most of the markets from/to Milan. This makes the AZ-LIN context an ideal setting for the IGASFAM to deliver significant improvements. In this research, we compare the current situation as of May 2018 with the profit-maximizing solution resulting from the application of the model. To highlight the advantages of the proposed approach, we consider two different scenarios in which (1) demand generation is fixed and demand allocation follows a simple multinomial logit model or (2) both demand generation and allocation are endogenized following the integrated nested logit model (IGASFAM).

3.5.1 Demand estimation

A key requirement for implementing the IGASFAM is to retrieve the coefficients of the demand model. Estimation data were collected from OAG, specifically, from two modules—the Schedule Analyzer and Traffic Analyzer[7] The resulting sample comprised itinerary-carrier observations on a monthly basis for medium-haul trips (distance $\leq$ 3,000 km) originating, transiting, or arriving in Italy in 2018. The total number of records was 127,100, corresponding to 4,592 directional city pairs. Due to

---

[6]Its capacity has been reduced from the technical capacity down to 18 scheduled commercial movements/hour, and only domestic or international flights within the EU are allowed (Airport regulation of Linate Airport (2016)).

[7]The OAG Schedule Analyzer provides data on schedules and capacity, while the OAG Traffic Analyzer provides historical data on airfares and passenger demand. Details on data sources can be found at the provider's official website, while details on how to assemble the data for estimation can be found in Birolini et al. (2020).

data limitations, we could not differentiate between passenger types in the following application[8]. Yet, the inclusion of market fixed-effects enables the estimation of different elastic response functions and substitution patterns to changes in air travel provision, thus capturing the specificities of each market in aggregate terms.

Consistent with the modeling framework, the saturated demand level in a given market must be independent of the features of the transport services provided and can be represented as a function of market demographic and socio-economic characteristics. The empirical estimation of a saturated demand function is challenging owing to the impossibility of observing saturated demand values. We followed the approach proposed by Hsiao and Hansen (2011) and computed the saturated demand in each market as the geometric mean between populations multiplied by a proportionality coefficient (α) that represents the maximum number of potential trips per capita. Let Pop_o and Pop_d be the population at the origin and destination of a given market, then $D = \alpha Pop_o Pop_d$. In this study, α was taken to be equal to the maximum number of trips per unit of population in the data sample[9] and, for each metropolitan area, the population was computed considering the number of residents within a radius of 50 km from its centroid[10]. To account for further factors that may affect trip generation (e.g., income level, tourism attractiveness), market fixed effects were added to the utility formulation of the air-travel nest (δ). In doing so, different demand patterns were acknowledged for various markets. In addition, the inclusive value (IV_{air}) was interacted with the distance between market metropolitan areas ($dist$) to proxy the intensity of competition with other transport modes.

At the itinerary level, the following supply-side variables were considered to proxy

---

[8]Although the OAG Traffic Analyzer module provides passenger data differentiated by fare classes, their use to estimate separate sets of coefficients for different passenger types is not straightforward. First, disaggregate data are largely missing; second, especially concerning inter-European travel, these data only reflects marginal variations in prices and do not suitably represent passengers' trip purpose. Third, data for fare classes are not homogeneous between FSCs and LCCs.

[9]This value, which is equal to 0.11, provides a reasonable and sufficiently large estimate to guarantee the consistency of lower-level coefficients. To ensure the robustness of the estimated parameters, sensitivity tests were performed for different values of α. Further details can be found in Hsiao (2008).

[10]Data were extrapolated from the GPWv4 dataset provided by the Socio-Economic Data and Applications Center (SEDAC) - Columbia University.

the itinerary deterministic utility component (V_i): flight time, connecting time, type of service (nonstop vs. connecting itinerary), and average price.

To estimate the model coefficients, we followed a sequential approach. In order to get consistent coefficients at the daily level, the following specification was used: $MS_i = \frac{N_i e^{V_i}}{\sum_{j \in I_m} N_j e^{V_j}}$, where MS_i is the market share and N_i the frequency of itinerary-carrier i. By taking the difference between logarithms of market shares of two alternatives serving the same market (e.g., i and j), we obtain the following linear equation, where X_k refers to the k-th itinerary attribute and β_k are the coefficients to be estimated:

$$\log MS_i - \log MS_j - \log(\frac{N_i}{N_j}) = \sum_k \beta_k (X_{ik} - X_{jk}) + \xi_{ij} \qquad (3.31)$$

Following Birolini et al. (2020), the estimation was carried out using a bootstrap procedure to jointly address issues of price endogeneity and airfare data missingness (28%). At each bootstrap repetition (1,000 in total), the average fare was first approximated using complete cases, including as relevant predictor the low-cost and time dummies, and the product of distance and jet fuel cost as a valid instrument (Hsiao and Hansen 2011, Gayle 2013); this first regression was used to provide instrumented prices for all observations in the bootstrap sample, which were then entered into the linearized itinerary choice regression in Equation (3.31).

Finally, the itinerary choice coefficients were used to calculate the inclusive values and estimate the upper level model coefficients involving the choice between the air travel nest and the outside good. Similar to Equation (3.31), we take the difference between the logarithms of the air travel market shares and the no-fly option ($V_{no-air} = 0$) thus obtaining the following linear function, estimated by OLS:

$$\log MS_{air} - \log MS_{no-air} = \delta + \theta IV_{air} + \rho IV_{air} dist + \xi \qquad (3.32)$$

Tables 3.2 and 3.3 list descriptive statistics for the variables included in the model and the estimation results. The coefficients are statistically significant and have plausible values. The inferred time values are 14.5 and 7.5 \$/h for flight and connecting

Table 3.2: Descriptive statistics

Variables	Average	Std dev	Data Range [25th, 50th, 75th]
Market share	28.70%	33.20%	[3.0%, 11.7%, 46.3%]
Flying time (minutes)	190.9	65.3	[145.0, 187.8, 230.0]
Connecting time (minutes)	111.7	84.3	[0, 120.0, 175.2]
Price (2018 US$)	155.6	105.1	[92.0, 133.0,187.0]
Population (million)	2.85	2.7	[1.29, 2.10, 4.13]
Market distance (1,000 km)	1.35	0.64	[0.87, 1.26, 1.72]

Table 3.3: Estimation results

Nest level		Variables	Coefficient	Std err
Demand Allocation (V_i)		FT - Flight time (minutes)	-0.0087***	0.000
		CT - Connecting time (minutes)	-0.0045***	0.000
		NS - Nonstop (0/1)	2.5665***	0.015
		P - Price (IV) (2018 US$)	-0.0359***	0.001
		R^2 adjusted	0.74	
		Number of observations	90,607	
Demand generation (V_{air})	δ	Constant	-4.204***	0.426
		Market fixed effects	YES	
	θ	IV–Inclusive value	0.571***	0.086
	ρ	IV*dist (1,000 km)	-0.088***	0.033
		R^2 adjusted	0.95	
		Number of observations	36,271	

Confidence levels: *$p < 0.10$. ** $p < 0.5$. *** $p < 0.01$.
Note: Standard errors are robust to heteroskedasticity, cluster and temporal correlations (Driscoll and Kraay' standard errors).

times, respectively, while the inferred premium for nonstop travel is 71.5 $. The nesting coefficient, which is lower than one and moderate in magnitude, indicates a significant degree of substitution between the air and non-air alternatives. Distance is seen to have a negative and relevant impact, confirming the logical intuition that more distant markets are characterized by lower competitive pressure from other travel modes. As a main drawback, since data are available on a monthly basis, we could not estimate significant time-of-day preferences, which were therefore tentatively ignored in the case study analyzed. Yet, to indirectly assess the impact of time-of-day preferences on model results, we conducted a sensitivity analysis replicating the time-of-day patterns considered in previous studies (see Section 3.5.4, for details).

3.5.2 Study data

To implement the IGASFAM, data on flight schedules, competition, passenger demand, average fares, and costs were collected from OAG, official reports, or obtained from Alitalia. Schedule-related information, such as flight and departure times, were assembled to compute the set of itineraries and related features for the target airline, partner, and competitor flights. We only considered nonstop and one-stop connecting itineraries. The baseline network consisted of 244 markets, of which 50 nonstop and 194 one-stop (Figure 3-5). Alitalia directly operates short- and medium-haul flights to European capitals and domestic destinations, which are characterized by high demand and intense competition. Table 3.4 reports statistics on the markets analyzed.

Figure 3-5: Alitalia's baseline network (May 2018).

The proportions of local and transfer demand through LIN were 83.7 and 9.7%, respectively, with beyond/behind traffic accounting for the remaining 6.6%. The

116

Table 3.4: Markets data summary

Markets	Variable	Average	Data Range [25th, 50th, 75th]
Nonstop	Number of competitor airlines	2.4	[2, 2, 3]
	Number of direct flights by competitors (daily)	7.6	[3.1, 5.5, 10.5]
	AZ market shares	26.00%	[12.1%, 17.7%, 35.1%]
	Market size (daily O&D pax)	1256.2	[477.1, 1164.8, 1846.8]
One-stop	Number of competitor airlines	1.7	[1, 1, 2]
	Number of direct flights by competitors (daily)	1.5	[0.3, 0.6, 1.7]
	AZ market shares	9.40%	[1.8%, 3.9%, 10.6%]
	Market size (daily O&D pax)	175.2	[14.8, 76.7, 144.5]

average number of daily flights operated was 180, with a minimum and maximum of 135 and 201 on Sundays and Wednesdays. Despite the focus of our analyses was placed on the Alitalia network based at LIN, we had to consider its interactions with the rest of the network. These interactions relate to Alitalia's own feeding flights onto the LIN-FCO route and the behind/beyond itineraries that are partially operated by partner carriers[11]. Following a similar approach to that used in Pita et al. (2013), the potential behind/beyond flight legs were considered to be fixed, and the model was allowed to allocate passengers over these flight legs as part of connecting itineraries. In the absence of detailed information on revenue sharing mechanisms, an equivalent average fare was computed based on the relative portion of travel distance operated by each airline.

Available flight options were validated with the company and mandatory flights were defined to ensure hourly frequencies on the LIN-FCO route and to meet the minimum flight requirements of the PSO routes to Sardinia; collectively, these flights accounted for about 32% of the total number of flights. Additionally, consistent with the regulatory requirements at LIN, slot constraints were enforced to ensure that the number of hourly movements at LIN did not exceed the current number of Alitalia's operations. The fleet considered was composed of four aircraft types—5 Airbus A320 with 171 seats, 14 Airbus A319 with 144 seats, 6 Embraer E190 with 100 seats, and 8 Embraer E175 with 88 seats—for a total of 33 aircraft the maximum allowable load

---

[11]Members of the same alliance (SkyTeam) or other carriers with which codeshare agreements are in place.

factor was set to 80%.

Direct operating costs include fuel, maintenance, and crew costs. Overhead and fixed costs, including administration and ownership costs, among others, were not considered as they do not affect the tactical decisions addressed in this study. Airport charges were collected from official documents. Airport charging schemes vary widely on a per-country basis, as they are regulated by national authorities. Therefore, we redirected them to two main components: airport charges that are applied per aircraft movement (landing and take-off fees, parking charge, and additional fees for the use of airport centralized infrastructures), and airport charges that are levied per passenger (boarding, passenger, and hold baggage security fees). Values for different aircraft types were computed based on their maximum take-off weights (MTOWs) and other technical specifications provided by the manufacturers.

3.5.3   Numerical results

Table 3.5 compares the main results of the application of the model to the baseline scenario. The baseline scenario refers to May 2018 considering the actual capacity allocation under the modeling assumptions and passenger predictions for comparability. Scenario 1 assumes a total market demand that is fixed at the baseline level, i.e., inelastic trip generation. Let $\bar{Q}$ be the total (fixed) market demand, then Equation 3.2 reduces to a basic MNL model, i.e., $q_i = \bar{Q}(eV_i/eV_{tot})$ and constraints 3.11 are adjusted accordingly. In doing so, we only endogenize demand allocation to replicate the approach used by previous studies (see Section 3.2). By constrast, Scenario 2 involves application of the IGASFAM with both demand generation and allocation considered simultaneously. For each scenario, we consider two different situations in which the total number of flights is constrained to be the same as the number of flights operated in the baseline scenario (a), or it can be arbitrarily lower (b).

Relative to the baseline scenario, the application of IGASFAM yields a significant increase in the operating profits up to 6.9% by reducing the number of flights by 4.4% and increasing the average revenue per passenger by 7.8% (Scenario 2b). The margin of improvement under scenario 1 is significantly lower (4.4%) and involves a stronger

Table 3.5: Summary of results for the four scenarios (daily values). Variation (%) compared to the baseline scenario in brackets.

Scenario		Operating profits	Revenues	Costs	Avg Fare	Tot Pax	Tot Flights	Aircraft Util
Inelastic market demand	*Scenario 1a*	(+0.4%)	(-3.9%)	(-5.9%)	(+3.8%)	13,364 (-7.4%)	180 (0%)	7.5 h (-5.6%)
	Scenario 1b	(+4.4%)	(-7.3%)	(-12.8%)	(+5.9%)	12,632 (-12.5%)	166 (-7.8%)	6.9 h (-12.2%)
IGASFAM	*Scenario 2a*	(+5.0%)	(-3.3%)	(-7.2%)	(+6.8%)	13,064 (-9.5%)	180 (0%)	7.4 h (-6.4%)
	Scenario 2b	(+6.9%)	(-5.4%)	(-11.3%)	(+7.8%)	12,654 (-12.3%)	172 (-4.4%)	7.1 h (-10.2%)

reduction of flight frequencies (-7.8%) (Scenario 1b).

The optimization model suggests increasing capacity on northern routes, such as Luxemburg, Brussels, and Dusseldorf, which are characterized by high potential demand, lower competition, and advantageous routing factors to connect with destinations in Southern Italy (Figure 3-6). Capacity can be retrieved by dropping frequencies to/from Madrid and main cities in Sicily (Palermo and Catania), where fierce competition mostly from LCCs tends to push down the revenues per available seat-kilometer and decrease profits. A comparison of Scenario 1 and 2 highlights the contribution of integrating demand generation and allocation. The results were similar in both scenarios in terms of routes for which frequencies should be dropped/enhanced but differed substantially in terms of the magnitude of network change. In particular, overlooking trip generation led to an optimal solution with a much higher number of unused slots, which is undesirable given their strategic value at LIN. This is further evident by comparing scenario 1a and 2a when the total number flights is fixed. While the application of the IGASFAM may still deliver substantial profit gains (+5.0% in Scenario 2a), enforcing all slots to be used under a fixed market demand scenario would leave little margin for schedule and fleet assignment optimization leading to profit gains of only +0.4% relative to the baseline scenario.

Without considering the contraction effect on total demand, the Scenario 1b results suggest cutting too many flights leveraging on the expectation of achieving higher load factors and improving capacity utilization. By contrast, under Scenario

(a) Scenario 1b (Inelastic market demand) **(b)** Scenario 2b (IGASFAM)
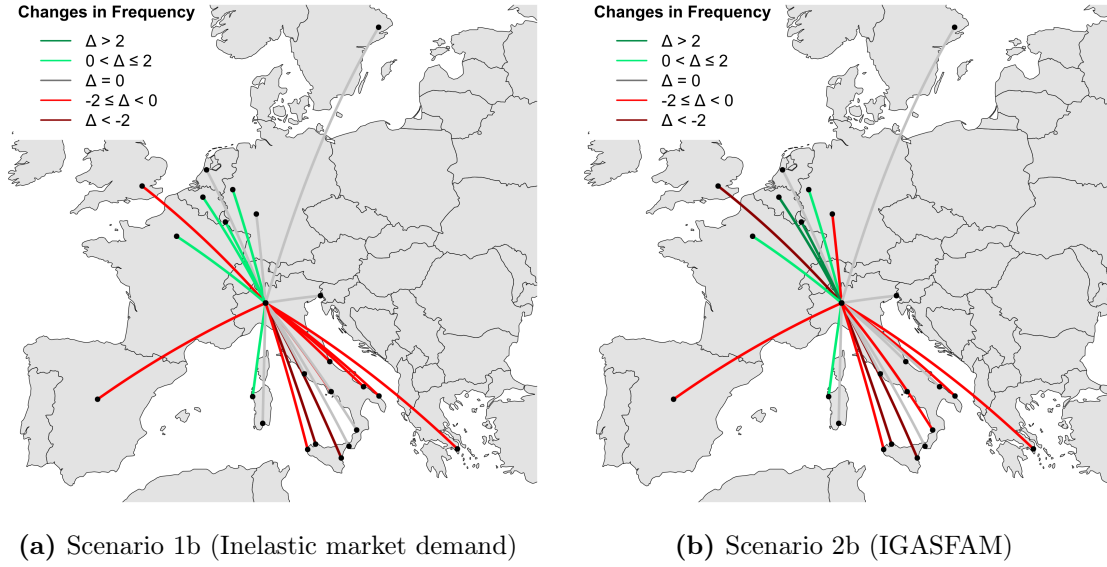
Figure 3-6: Changes in flight frequencies by markets compared to the baseline scenario.

2b, the two effects that follow from the deletion of itineraries – the spilling of demand to other itineraries and a reduction in total demand – appear to be better traded-off. To further emphasise this aspect, Figure 3-7 depicts the patterns of change (in percentage terms with respect to the baseline) in the level of supply, total market demand, and itinerary demand across nonstop markets. Under Scenario 1, market demand is fixed and inelastic; therefore, itinerary demand tends to be overestimated when supply is reduced, while underestimated when supply is increased. By contract, under IGASFAM, total market demand correlates to the level of supply, suitably capturing demand contraction and stimulation effects, and leading to more reasonable variations of q_i. Consistently, a closer look into the solutions obtained reveals that, under the full implementation of the IGASFAM (Scenario 2), better opportunities for demand stimulation in underserved markets can be identified. Under the fixed-demand scenario, by contrast, the model tends to allocate capacity in large and congested markets, leading to an optimal solution that more closely resembles that of the baseline scenario. Collectively, these results highlight the practical benefits of explicitly considering demand generation and allocation dynamics into flight scheduling and fleet assignment models.

(a) Scenario 1b (Inelastic market demand)   **(b)** Scenario 2b (IGASFAM)
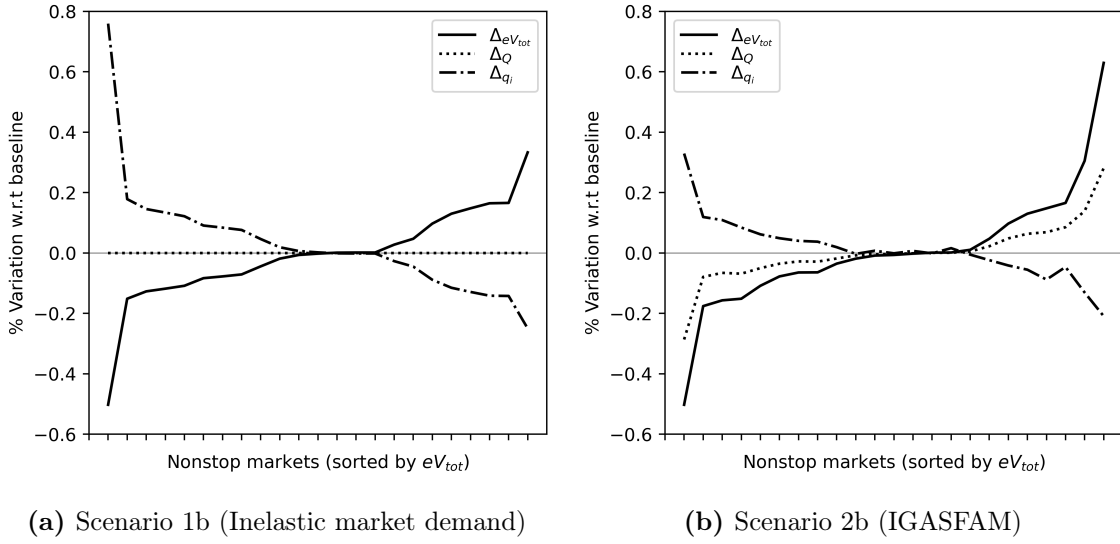
Figure 3-7: Changes in the level of supply (proxied by eV_{tot}), total market demand Q and itinerary demand q_i.

3.5.4   Sensitivity analysis

Finally, to assess the robustness of the model results, sensitivity analyses were performed on key model parameters and assumptions. First, Table 3.6 summarizes the computational performance and changes in the optimal solution for a different number of piecewise segments (1, 2, and 3), which represents a key assumption of the IGASFAM. While increasing the number of breakpoints reduces the approximation error, it can lead to a larger instance size and consequently a longer computation time. Hence, one would like to identify a number of pieces that suitably trades off the solution quality and computational cost of the model. If only one line segment is used to approximate the demand function (Equation 3.11), the number of passengers is strongly overestimated (+18,80%) due to the convex nature of the itinerary demand curve, thus yielding unrealistic and overconfident results (+50.23% increase in operating profits compared to the base value, i.e., two pieces). On the contrary, using three pieces leads to the exact solution obtained with two pieces in terms of allocated capacity. Minor (downward) variations are brought to the selected financial indicators (+0.44% profits), but the small gain in accuracy comes at the expense of a much greater computational cost, with the two-pieces approximation taking about

121

65 minutes and the three-pieces more than 20h (optimality gap equal to 0.66%) to solve to optimality. Therefore, the choice of using two linear pieces appears as a valid and suitable approach for the practical implementation of the IGASFAM.

Table 3.6: Sensitivity analysis for the number of linear pieces referred to the application of IGASFAM (Scenario 2b).

Nr. Segments	Operating margin	Revenues	Costs	Tot Pax	Computation time (s)	Accuracy-R^2 [min, mean, max]
1	+50.23%	+19.42%	+1.85%	15,033	89	[0.31, 0.93, 0.99]
2 (base value)	-	-	-	12,654	3,917	[0.94, 0.98, 0.99]
3	-0.44%	-0.70%	-0.84%	12,570	>72,000	[0.98, 0.99, 0.99]

Second, Table 3.7 summarizes the results of the sensitivity analyses with respect to four aspects — time-of day preferences, competition, saturated demand, and maximum allowable load factor. Time-of-day preferences were modeled assuming time-of-day patterns empirically estimated and used in previous studies. Specifically, we consider the approaches followed by Atasoy et al. (2014) and Pita et al. (2014), where the former implies a higher utility for itineraries departing in the morning (between 7 and 11 am), and the second considers two peak periods in the morning (arrival time between 8 and 10 am) and in the evening (departure time between 6 and 8 pm). The benefit of traveling during peak periods was assumed to increase the itineraries' exponential utilities by either 10% or 5%. Results reveal that leveraging time-of-day preferences could lead to a better optimization of flight departures throughout the day. Yet, the overall impact on operating profits (ranging between 1.15% and 2.85%) is limited, thus practically corroborating the IGASFAM tactical insights.

The impact of competition was tested by varying the eV_{fix} parameters. Reasonably, an increase in the level of competition negatively affects airline's economic performances (-8.22% and -4.21% in operating profits upfront +10% and +5% in competition, respectively), while a decrease in the competitive pressure results in higher market shares, which, in turn, enable a significant increase in revenues and profits (+8.53% and +4.23% in operating profits upfront -10% and -5% in competition, respectively). Increases in economic performances could also be driven by demand growth. We observe an almost linear improvement of 9.48% (4.85%) in revenues

when saturated demand increases by 10% (5%) which, along with unvarying costs, determines significant profit surges (25.22% and 12.86%, respectively). Note, however, that when relevant changes affect the saturated demand values, a re-estimation of model coefficients is required to ensure adherence with respect to the normalized "no-fly option". Finally, we investigated the impact of variations in the maximum allowable load factor. An increase in this parameter leads to marginal performance improvements (2.24% and 1.42% in operating profits for a 100% and 90% maximum allowable load factor, respectively). By contrast, a decrease in this threshold (below average load factors in practice) would significantly affect the estimated profits (-3.63% and -11.80% for maximum allowable load factors of 70% and 60%, respectively). Collectively, these analyses corroborate the validity of the proposed modeling framework.

Table 3.7: Sensitivity analyses to the application of IGASFAM (Scenario 2b).

Parameter	Variation	Op. profits	Revenues	Costs	Avg Fare	Tot Pax	Tot flights	Avg Aircraft Util
Time of day	10%	2.68%	0.59%	-0.61%	0.15%	0.44%	0.00%	-0.55%
(morn. peak)	5%	1.45%	0.53%	0.01%	-0.01%	0.54%	0.00%	0.00%
Time of day	10%	2.85%	1.14%	0.16%	-0.08%	1.21%	0.00%	0.00%
(morn./even. peaks)	5%	1.15%	0.61%	0.30%	-0.04%	0.64%	0.00%	0.00%
	-10%	8.53%	3.22%	0.19%	0.09%	3.13%	0.00%	0.00%
	-5%	4.23%	1.66%	0.19%	0.00%	1.66%	0.00%	0.00%
Competition	5%	-4.21%	-1.54%	-0.02%	-0.04%	-1.51%	0.00%	0.00%
	10%	-8.22%	-3.21%	-0.36%	-0.07%	-3.14%	0.00%	0.00%
	-10%	-26.80%	-11.10%	-2.15%	0.42%	-11.47%	-1.16%	-1.39%
	-5%	-13.61%	-5.66%	-1.12%	-0.08%	-5.58%	-1.16%	-0.77%
Saturated demand	5%	12.86%	4.85%	0.28%	0.14%	4.71%	0.00%	0.00%
	10%	25.22%	9.48%	0.50%	0.05%	9.43%	0.00%	-0.18%
Maximum	-20 pp	-11.80%	-4.90%	-0.96%	3.98%	-8.53%	-1.16%	-3.32%
allowable load	-10 pp	-3.63%	-2.04%	-1.13%	0.40%	-2.43%	-1.16%	-1.69%
factor (base value:	+10 pp	1.42%	0.14%	-0.58%	-0.04%	0.18%	0.00%	0.00%
80%)	+20 pp	2.24%	0.35%	-0.72%	0.03%	0.33%	0.00%	0.16%

3.6 Conclusion

In addressing the supply-demand interactions in airline planning, this paper contributes to the literature by describing a new formulation to endogenize demand

within itinerary-based integrated flight scheduling and fleet assignment models in a utility-consistent manner.

To model demand in an integrated fashion, a nested logit model that includes the non-air travel alternative is used, which provides sensible substitution patterns and solves the demand generation and allocation problems simultaneously. The incorporation of the nested logit formulation leads to a nonlinear mixed-integer formulation that can be effectively solved by introducing tightening constraints and using a tailored piecewise linearization scheme.

To illustrate the validity of the proposed approach, two separate analyses were conducted. First, the computational tractability of the model was investigated through computational experiments carried out using randomly-generated realistic problem instances. Comparison of the results with real-world cases highlighted the tractability of the model and demonstrated how it can properly handle midsize networks. Second, to demonstrate the managerial insights that can be derived from the application of the model, it was implemented in a real case study using data on the domestic and international network of Alitalia through the Milan-Linate airport. A comparison of data obtained for May 2018 with profit-maximizing solutions obtained through the application of the IGASFAM highlighted the significant improvements that the model can deliver relative to the baseline scenario (an increase in operating margin of up to 6.9%). The specific contribution of integrating demand generation and allocation was investigated by comparing the model's optimal solution with that produced by an inelastic demand-generation scenario. In this case, the expected improvement was +4.9% and, more importantly, the IGASFAM was shown to be better at revealing opportunities for demand stimulation in less-served markets. Thus, despite the necessary simplifications in the model, the results indicate that the model can effectively support airlines' decision-making.

Extensions of the IGASFAM model can provide further insights into the improvement of airline planning processes. One potential development would relate to the use of the model for a systematic assessment of route development alternatives and support entry decisions. In this regard, the model might be able to deal with additional

complexities that arise from the strategic reaction of competing airlines. A potential approach would be to solve the IGASFAM model iteratively to explore the competitive dynamics in a game-theoretical framework. From a computational viewpoint, the nonlinearities and complex interactions in the model make the development of heuristic methods an interesting research direction for extending its applicability to large-case scenarios. Eventually, research could be carried out to improve the utility specification and evaluate how the consideration of additional itinerary attributes, such as time-of-day preferences, might affect the optimal solution.

Chapter 4

Integrated Origin-based Demand Modeling for Air Transportation

4.1 Introduction

The estimation of travel demand is an important priority in the air transport industry and a key issue in academic research. Demand modeling in an air transport context poses important challenges arising from its intrinsic unpredictability and the presence of actors with diverse interests. From the longest-term strategic decisions to daily schedules, airlines require reliable forecasting to make optimal decisions at all stages of the planning process (Lohatepanont and Barnhart 2004, Carreira et al. 2017), while projections of traffic flows are crucial for planning large and irreversible investments in the expansion, renovation, and maintenance of terminal and flight infrastructures (Flyvbjerg et al. 2005, Xiao et al. 2016). Preventively understanding the patterns of passenger growth is also key for policymakers to establish strategic transportation plans for the future development of national and international connectivity (Park and Ha 2006, Hakim and Merkert 2016).

The prevailing approach in the demand forecasting literature has been to consider city pairs as the most appropriate definition of air transport markets (Brueckner et al. 2014). While this is motivated by the gravity-like nature of socioeconomic interactions that originate air travel between origin and destination areas, the estimation

of air traffic flows in a city pair should not disregard the potential interdependences with other markets. Taking the Milan-New York city-pair as an example, if a new destination is served from Milan, e.g., Washington, it is reasonable to expect that some passengers on the Milan-New York route will instead choose to travel to Washington, to the extent that there is a similarity of destination and travel purpose. Adding Washington as a destination might also induce more passengers to fly. Fundamentally, this reasoning applies to any change in the air transport supply, ranging from minor variations in the attributes of existing itineraries to the introduction of brand-new destinations.

The rationale to account for alternative destinations in the estimation of origin and destination (O&D) flows draws from the classical transport literature (Ortuzar and Willumsen 2011), according to which each region has a potential for generating passenger trips (i.e., trip generation) that are allocated among available destinations in proportion to their respective attractiveness (i.e., trip distribution). However, in the air transport industry saturated demand is typically assumed, either implicitly or explicitly, not to occur at origin zones but rather independently within city-pairs. The main drawback of this approach is that it overlooks potential substitutions between markets and, more importantly, can lead to over- or under-estimation of trip-end totals, i.e., the total amount of trips generated from a given region.

Drawing upon the contributions of Wei and Hansen (2005) and Hsiao and Hansen (2011), who advanced the assessment of air passenger demand by integrating trip generation and allocation at the city pair level, in this paper we offer an aggregate origin-based demand model that assumes trip generation at the origin level and explicitly accounts for the degree of substitution between destinations. The proposed model employs a multilevel nested logit (NL) approach that starts from the choice as to whether or not to travel by air. Branching down the NL tree, we group possible destinations by trip length and further decompose each city pair into their inner travel alternatives, i.e., air travel itineraries. The use of an NL formulation allows for the correlations between alternatives to be captured at different choice levels, the simultaneous solution of different stages, and the highlighting of travelers' behavioral

responses to changes in air travel provisions at both the itinerary and destination levels. Ultimately, we propose a bootstrap procedure to cope with missing data and price endogeneity issues within the aggregate logit estimation. The proposed integrated model was validated through testing over the comprehensive network of outbound air trips from Italian airports in 2018 using airfare and passenger data retrieved from the Official Airline Guide (OAG).

The remainder of this paper is organized into eight sections. In Section 4.2, we explore the theoretical background of the problem through relevant literature and specify the contribution of this work. Section 4.3 describes the modeling framework, Section 4.4 focuses on the methodology, and Section 4.5 presents model estimation issues. In Section 4.6, we present the empirical setting and in Section 4.7 we present the robustness checks and report our results. Finally, in Section 4.8 we conclude the paper and discuss directions for future research.

4.2 Theoretical background

In previous studies, air travel demand has primarily been analyzed by dividing the problem into two main sub-problems: demand generation and demand allocation. Demand generation refers to the estimation of total air traffic flows while demand allocation involves the study of factors underpinning the distribution of passengers over available travel itineraries within a specific market.

Air demand generation is typically modeled via a gravity formulation approach (Wojahn 2001, de Grange et al. 2010, Adler et al. 2018) in which travel demand between city pairs is assumed to be positively proportional to the mutual attraction factors of the respective cities and inversely proportional to the generalized cost of travel between them. Different cost and incentive factors have been used to evaluate what impedes, or facilitates, air traffic flows between territories. In general, attractor (pull) factors depend on regional, geographical, and socio-economic characteristics (demand-side factors) such as income and population (see Grosche et al. (2007) for a review), whereas supply-side (push) factors are used to model the ease or difficulty

of traveling between regions. In its simplest form, travel impedance is modeled as the distance between considered regions and advanced formulations include specific characteristics of transportation systems to account for their role in enabling travel. In this regard, numerous supply-side variables have been considered, primarily in an aggregate fashion, including frequency, travel time, average price, and the availability of low-cost carriers (LCCs) (e.g., Jorge-Calderón 1997, Boonekamp et al. 2018). Although gravity models have been extensively used, they have a major limitation in that they assume independence between origin-destination pairs. Another class of trip generation models is represented by multiplicative models (Belobaba 2009), which differ from a gravity formulation in their functional form but share similar limitations concerning the independent treatment of city-pair demand. As demonstrated by Margaretic et al. (2017), overlooking the spatial dependence between air transport markets is a relevant issue as it may yield unreliable parameter estimates. The importance of accounting for correlation between trips that originate from the same origin zone is further highlighted in de Grange et al. (2011), who proposed a hierarchical gravity model to accommodate various spatial correlation structures.

Along with the study of demand generation, several passenger allocation models have also been developed. An important class of allocation models is the so-called itinerary choice models, which focus on how individuals select among air-travel alternatives within a city-pair market. These models have investigated the role of a variety of determinants, highlighting the importance of level-of-service, connection quality and carrier attributes (Coldren et al. 2003) (e.g., airfare, travel and connecting time, number of stops, and frequency), punctuality (Freund-Feinstein and Bekhor 2017) and time-of-day preference (Koppelman et al. 2008, Lurkin et al. 2017). Specific aspects of demand allocation have been further examined concerning not only the features of air travel itineraries but also the presence of alternative transport modes (Adler et al. 2010), the competition between neighboring airports (e.g., Başar and Bhat 2004, Hess and Polak 2005), and the ground access and egress portions of the air passenger trip (Pels et al. 2003).

The aggregation of trip generation and demand allocation modeling has received

increased attention recently as a method for better capturing the interactions between air transport demand and supply. Although free-standing demand generation models are known to perform well in estimating macroscopic traffic flows between territories, they do not fully appraise the specificities of single itineraries and how changes in their attributes can impact the overall air traffic volume. On the other hand, demand allocation models provide a valuable tool for estimating itinerary market shares, though in most practical applications they rely on a static representation of air total demand. Departing from these limitations, a few contributions to the air transportation literature have analyzed both demand generation and allocation in tandem. (Wei and Hansen 2005) and (Hsiao and Hansen 2011) proposed similar approaches to tackling the two stages simultaneously based on aggregate NL models. In the first study, a two-level NL model was developed to investigate the determinants of airline market share and total air travel demand in a non-stop duopoly market; in the second, a general three-level NL model was proposed to simultaneously estimate the overall air travel demand in a city pair and distribute it over the available itineraries. The greatest advantage of such approaches to integrated modeling is that they make demand generation elastic by dynamically incorporating the compound utility provided by air travel itineraries at the trip generation stage.

To the best of our knowledge, previous studies that estimate demand on air travel itineraries have considered each city-pair market separately, with air trip generation occurring at the city-pair level and independently of other markets. This formulation is restrictive in that assumes that air travel between two cities is not affected by changes in air transport supply elsewhere. Although this might be a safe assumption for separate and disjointed markets, i.e., cases in which neither the origin nor the destination is common (See Figure 4-1), it is likely not between overlapping ones.

In this context, this paper presents an integrated, origin-based air travel demand model that assumes saturation of demand at the territorial level. This allows us to properly consider the originating travel forces by area and suitably capture the substitution patterns between alternative destinations for originating outbound passengers. Within this framework, air trips are segmented according to the length of

131

Figure 4-1: Overlapping and independence amongst air transport markets (city-pairs).

Each circle in 4-1 represents an independent catchment area; i.e. multi-airport regions. Hereafter, we refer to overlapping markets as city-pair markets with the same origin or destination. Unlike commuter trips or intercity travel, in the case of air transportation we do not assume that a fixed saturation level exists for the number of trips attracted by each destination (e.g. as a function of jobs in the area). However, return trips on O-D and O-D' (Fig. 1c) are determined by the prior choice of people living in D to travel to O and O', respectively, which, among other things, strongly depends upon the availability of destination markets from D and the available transport service.

haul with the goal of accounting for differences in passengers' choice behavior and the travel determinants that affect long- vs. short-haul travel. Along with other issues (e.g., period of stay, overall cost of travel, etc.), the value that passengers place on air-travel-related features (e.g., travel time, travel cost, frequency, number of connections, in-flight services) is likely to vary with the length of haul, especially when comparing very long-distance trips with short/medium-haul ones. In this respect, separate modeling yields a better understanding of the dynamics of trip generation and allocation and provides insights into their elasticities and sensitivities to parameters of trips with different haul lengths. Previous studies have examined, for example, the differences between long- and short-haul trips in terms of passenger sensitivity to changes in airfare (Crouch 1994), connecting flights (Francis et al. 2007), and time of day (Lurkin et al. 2017). However, the study of how modifications to itinerary attributes can impact air trip generation and demand stimulation, and the substitutability between destinations at different haul lengths has remained a largely unexplored topic.

Building on these gaps, this paper presents the first systematic attempt to apply an origin-based approach to integrated demand generation and allocation modeling. This contributes to advancing the literature on air travel demand estimation by simultaneously considering the allocation of passengers over itineraries and alternative destinations, while constraining air trip generation on a regional basis.

4.3   Modeling framework

In this section, we describe the notation and theoretical considerations underpinning the modeling framework designed to represent our integrated origin-based air travel demand model. We denote the sets of origins and destinations as O and D to represent the geographical territories from which trips are generated and to which trips are attracted, respectively. Origins and destinations are not restricted to single airports but represent urban agglomerations closely bound to an attractor center by commerce or tourism, also referred to as metropolitan areas[1]. The total number of trips that an originating area can generate is constrained by its number of residents along with their propensity to fly (number of air trips per capita) as a function of regional socio-economic characteristics and the overall level of air service supply.

For each origin o, $D_o \subset D$, indexed by d, is defined as the subset of air-reachable destinations. As discussed in Section 4.2, not all destinations are evaluated homogenously as comparable alternatives for possible travel in any choice situation. For instance, it is unlikely for a passenger from Milan to consider a two-hour flight to London as an alternative on the same footing as traveling to New York for a weekend getaway (Figure 4-2). In addition to individual considerations and city-specific features, these two destinations are generally not perceived by passengers from Milan

---

[1]A number of contributions in the literature have focused on the proper identification of origin and destination areas in defining air transport markets (Brueckner et al. 2014). In place of a single airport's catchment area, these formulations attempt to represent multi-airport regions in which airports with competitive interactions are grouped. For example, as shown in Figure 4-2, the Milan metropolitan area is served by three main airports—Milan-Malpensa (MXP), Milan-Linate (LIN), and Milan-Bergamo (BGY)—which compete for local passengers and provide access to the region for incoming passengers. In this work, we assume there are no spatial interactions between metropolitan areas and we consider for simplicity a fixed "as the crow flies" radius around each zonal centroid.

as direct competitors: while New York can be seen as a closer substitute to Tokyo (ceteris paribus), London is likely seen as a more similar destination to Paris or any other large city with similar characteristics that could be reached by incurring a comparable generalized cost of travel. To account for systematic differences in preference between flight lengths, destinations can be further categorized into H lengths of haul, e.g. long-, medium-, and short-haul[2], and treated separately.
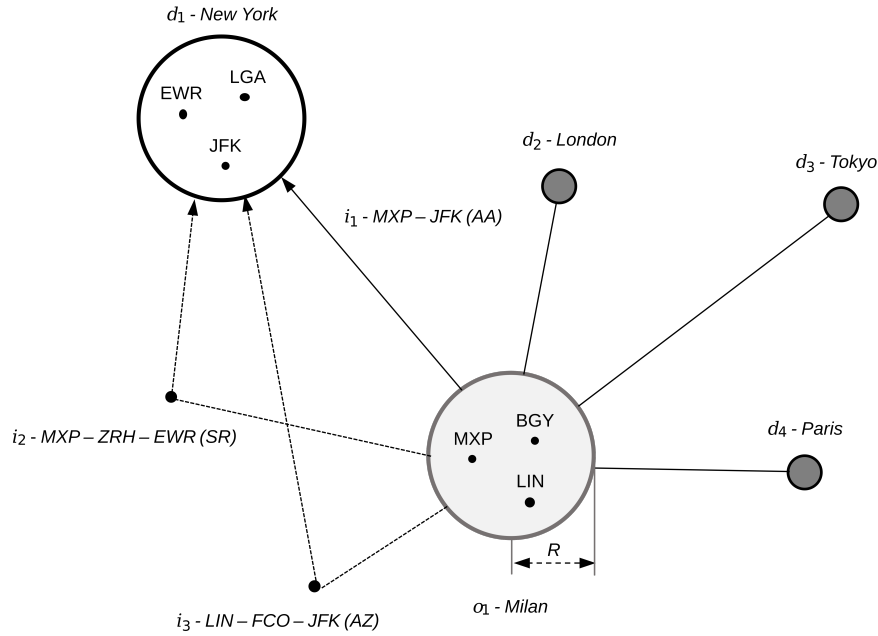

Figure 4-2: Origin-based framework.

Within a single city pair, e.g., Milan–New York (Figure 4-2), air transport supply can be well-diversified, offering passengers numerous choices of alternative routes. The set of air travel itineraries between an origin o and destination d, indicated by $I_{od} \subset I$, constitutes the elemental alternatives in the choice set available to passengers. Itineraries represent feasible travel options in a given market, can be nonstop or

---

[2]There is no international standard for the definition of route categories based on travel length. Commercial flights tend to rely on the conventional three-way classification into long-, medium-, and short-haul flights, although geographical differences lead to different cut-off conditions. In Europe, Eurocontrol defines short-, medium-, and long-haul routes as shorter than 1,500 km, between 1,500 and 4,000 km, and longer than 4,000 km, respectively [Source: Eurocontrol]. Other sources propose instead a classification based on the non-stop flight time; for instance, IATA [Global Passenger Survey (GPS)] distinguishes between short-/medium- (less than 5 h flight time) and long-haul routes (5 h or more). The proposed approach is applicable irrespective of the specific classification used as long as it can adequately reflect different preferences in travel.

multi-step, and are uniquely identified by the combination of airports, including the departure and arrival airports as well as any potential intermediate transfer hubs, and air carriers operating the flights. For example, the itinerary *LIN–FCO–JFK (AZ)* in Figure 4-2 represents the single connecting itinerary from Milan Linate (LIN) to New York John F. Kennedy (JFK) through Rome Fiumicino (FCO) operated by Alitalia (AZ), while *MXP–JFK (AA)* represents a competing nonstop itinerary provided by American Airlines (AA) from Milan Malpensa (MXP).

The proposed modeling approach estimates the originating outbound air traffic flows on each itinerary (Q_i) as a share of the maximum potential trips generated by a given area. We define T_o, the saturated demand at origin o, as the maximum number of trips that can be generated from that area; T_o includes trips that are currently made either by air or other transport modes and others that could be made but are not as a result of existing travel impedance. Assuming that T_o is known, the originating outbound air traffic flows on itinerary i are given as:

$$Q_i = T_o P_i \tag{4.1}$$

where P_i models the simultaneous choice to travel by air to a given destination using a specific itinerary i at an aggregate level, representing the absolute market share of itinerary i over the entire set of alternatives from origin o. P_i is assumed to be a function of a number of itinerary-specific service level characteristics, such as air fare and flying and connecting time, as well as destination-specific attributes (see 4.5.2). Equation 4.1 resembles the formulation proposed by Hsiao and Hansen (2011), differing primarily in that it defines the saturated demand uniquely at the territorial level rather than for each city pair. In our origin-based framework, this formulation replaces the origin-destination flow generation principle with an origin-based conservation of flow constraint. As the number of trips between city pairs is not independently determined, this modeling approach has the advantage of explicitly considering the degree of correlation between alternative destinations, such that stimulation of air travel demand in and out of a region can be achieved by improving the air transport supply,

both through incremental modifications to the attributes of existing alternatives and, more importantly, the development of new itineraries and destinations.

As in Li and Wan (2019), this study takes a directional approach to estimating the outbound trips originating from each origin area, herein referred to as originating outbound trips. Given that consumers in an air transport market typically start their trips in the origination region (the outbound portion of a passenger air trip) and return there after a trip of varying duration (the inbound portion of a passenger air trip), returning inbound flows may also be approximated by the modeling framework proposed in this paper. In addition, the supply of air service in each air travel market is shared with the respective opposite market that consists of passengers who originate their trips from the destination region. Two different models can be used to addressed this issue, one centered on the origin region and the other centered on the destination region. Consider for instance the Milan-New York city pair shown in Figure 4-3. The portion of originating outbound trips, representing the passengers originating from Milan in their outbound trip (MIL-NYC), is estimated using the Milan-centered model while the number of passengers that are returning to New York (MIL-NYC) can be approximated using a New York-centered model. The inverse is also true in relation to trips originated in New-York (NYC-MIL) and returning trips to Milan (NYC-MIL).

4.4 Methodology

To practically estimate the origin-based air travel demand model while accounting for differences in lengths of haul, we propose two different aggregate NL formulations[3].

Formulation 1 (NL1): The first formulation (Figure 4-4a) is a four-level NL model in which different lengths of haul represent the first level of nesting following the

---

[3]The NL specification has some practical advantages in estimating aggregate supply-demand models. First, it accommodates different correlation patterns within and between nests, making explicit the degree to which the utility obtained from the lower-level choice situation influences the upper model Train (2003). Second, the NL approach is based on the assumption of simultaneous choice; despite its tree structure representation, the NL model does not imply any sequentiality in the choice process, thereby avoiding the problem of inelastic trip generation. Third, NL models can be calibrated consistently using aggregate data, as discussed in Section 4.5.1
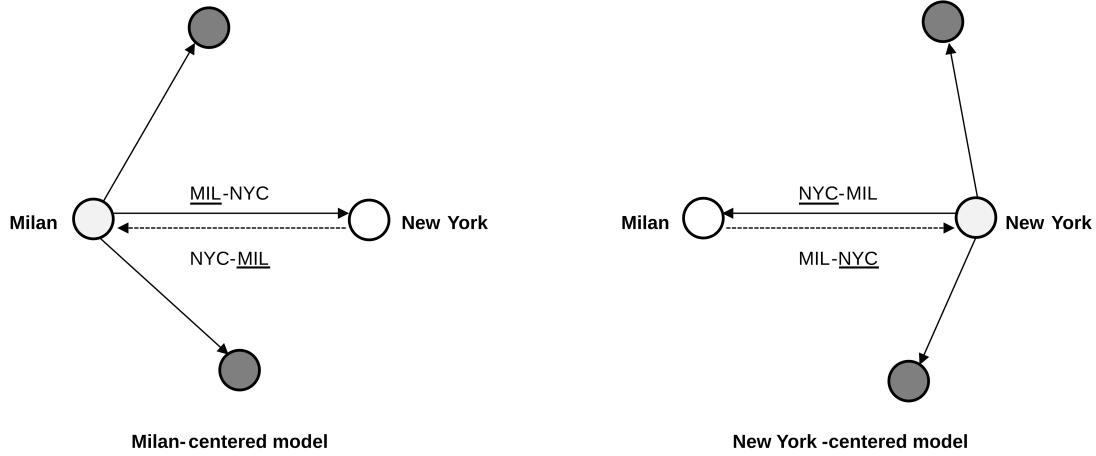
Figure 4-3: Representation of outbound and inbound traffic flows in a city-pair.

Solid arrows represent originating outbound traffic flows whereas dotted arrows stand for returning inbound traffic flows related to passengers who originated their trip in the model's center region (underlined city code).

decision to travel by *air*. This implies that trip making is not independent of length of haul and allows for changes in the attributes of an alternative to affect, to some extent, the overall travel demand on other haul lengths. In this respect, the nested structure accommodates flexible patterns of substitution that are expected to exist within and between haul length categories. Indicating the choice to fly as *air*, the integrated travel decision-making can be generalized as the following probability choice system:

$$P_i = P_{i|d,h,air} P_{d|h,air} P_{h|air} P_{air} \tag{4.2}$$

Formulation 1 (NL2): The second formulation (Figure 4-4b) assumes that trips over different lengths of haul exhibit significantly different features and therefore need to be modeled independently. As pointed out in Section 4.2, factors affecting the inertia and nature of travel can differ widely according to the scope of travel, particularly between short-haul and long-haul routes, eventually resulting in different propensities to fly and associated demand elasticities. Under this formulation, a separate three-level nested structure is estimated for each length of haul, with the choice to fly further partitioned into destination and itinerary nests. Indicating the

choice to travel by air over the haul length category h as air_h, the integrated trip decision-making process can be formulated as follows:

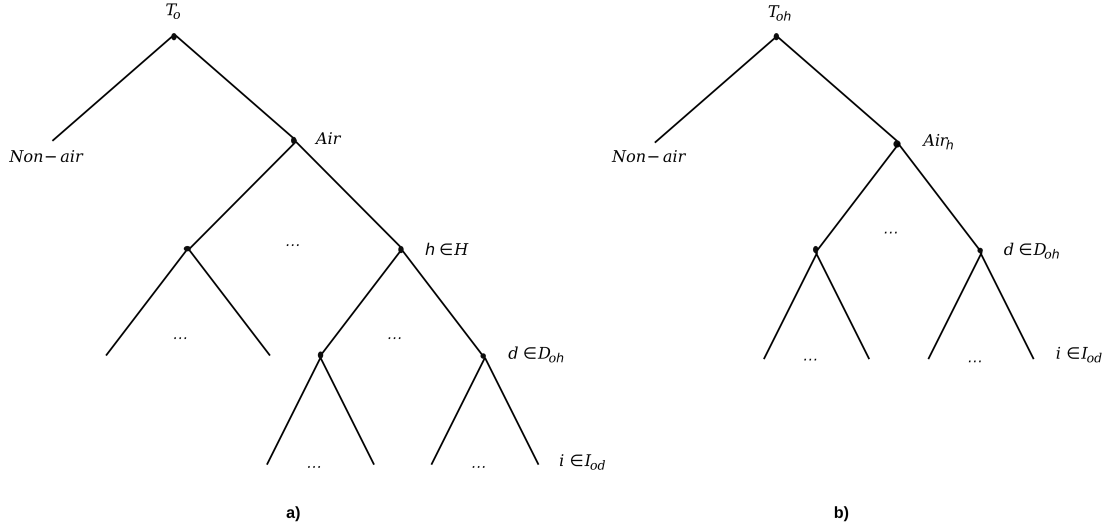$$P_i = P_{i|d,air_h} P_{d|air_h} P_{air_h} \qquad (4.3)$$

Figure 4-4: Nesting structures. (a). NL1, (b). NL2.

Both formulations imply hierarchical grouping of air travel itineraries by mode first (air/non-air) and then by destination as in Furuichi and Koppelman (1994). This nesting structure is practically convenient for estimating the demand for air travel in that, lacking specific data on competing modes of transport in each O&D pair (which is a common issue in practice), it allows for the normalization of a unique outside good at the root level and the estimation of consistent substitution patterns among air transport alternatives. The choice of the most appropriate formulation should be guided by theoretical considerations based on the given situation and confirmed by the empirical consistency of the nested structure with respect to utility maximization theory (Train 2003, Gil-Moltó and Hole 2004). The assumption of trip generation independence appears to be more or less restrictive according to the haul length classification being used. Other relevant factors are the travel purpose, which directly affects the propensity to consider alternative destinations, and the journey type, e.g., weekend getaway vs. two-week holiday, which inevitably affects the inclination to

look at faraway and more commoditized short-haul destinations as alternatives on the same footing.

Tables 4.1 and 4.2 contain the general formulation of the four-level nested logit model (NL1)[4]. At each level of the tree, the corresponding conditional probability takes the form of a multinomial logit formulation among nests. The deterministic utility component of each nest V) can be divided into two parts: a vector of specific explanatory variables that differ between nests but are common to all alternatives within the same nest; and the so-called inclusive value (IV), which incorporates the composite utility derived from the lower branches. The formulation in Table 4.2 originates from the adoption of the Random Utility Model 2 (RU2) normalization proposed by Hensher and Greene (2002)[5]. For the purpose of identification, scale parameters are normalized at the upper level and are restricted to constant values within each nest at the itinerary level. Unrestricted coefficients can be estimated at the destination level, but as this would result in a different set of estimated co-efficients for each destination, complicating the generalization and interpretation of model results, different scale parameters are only estimated across the haul length nodes to capture differences in substitution and contribution to overall travel demand by flights on different hauls.

---

[4]Although referred to as NL1, these equations can be applied to NL2 by dropping the haul length level and estimating separate models for each haul length category. Additionally, the model may be extended to allow for more flexible substitution patterns among itineraries by further grouping them based on, for example, the time-of-day (Coldren and Koppelman 2005), departure, arrival, or hub airport (e.g., Hsiao and Hansen 2011), and carrier type (LCCs vs FSCs) (Cadarso et al. 2017). Because the main objective of this paper is to present the upper part of the modelling framework—from the saturated demand to itineraries, which already entails three (NL2) to four (NL1) nesting levels—to demonstrate the usefulness of adopting an origin-based approach, we do not focus specifically on how to further branch down the itinerary nests and therefore tentatively ignore any further correlation patterns that might exist among them.

[5]There has been significant debate in the literature on the normalization and equivalence of NL models. It is widely acknowledged that RU2, which is equivalent to the utility maximizing nested logit (UMNL) specification (Koppelman and Wen 1998) is preferable to RU1/NNNL because it is consistent with utility maximization behavior in general cases. In particular, RU2 is able to provide consistent probability estimates for general model formulations involving unrestricted scale parameters across nodes on the same level of a tree or in the presence of generic attribute parameters (for a detailed overview, see Carrasco and de Dios Ortúzar (2002).

Table 4.1: Model notation (NL1)

Level	Parameter vectors	Specific explanatory variables	Scale parameters
Itineraries	β	X_i	
Destinations	γ	X_d	$\mu_d = \mu_{d'} = \mu \quad \forall d, d' \in D_{oh}$
Haul lengths	δ	X_h	$\lambda_h \neq \lambda_{h'} \quad \forall h, h' \in H$
Air/non–air	ω	$X_{air}, X_{non-air}$	

Table 4.2: Model formulation (NL1) - Conditional probabilities and utility functions.

Level	Utility	Probability	Inclusive Value
Itineraries	$V_{i|d,h,air} = \beta X_i$	$P_{i|d,h,air} = \dfrac{e^{\mu V_{i|d,h,air}}}{\sum_{i' \in I_{od}} e^{\mu V_{i'|d,h,air}}}$	$IV_d = \log\left(\sum_{i' \in I_{od}} e^{\mu V_{i'|d,h,air}}\right)$
Destinations	$V_{d|h,air} = \gamma X_d + \frac{1}{\mu} IV_d$	$P_{d|h,air} = \dfrac{e^{\lambda_h V_{d|h,air}}}{\sum_{d' \in D_{oh}} e^{\lambda_h V_{d'|h,air}}}$	$IV_h = \log\left(\sum_{d' \in D_{oh}} e^{\lambda_h V_{d'|h,air}}\right)$
Haul lengths	$V_{h|air} = \delta X_h + \frac{1}{\lambda_h} IV_h$	$P_{h|air} = \dfrac{e^{\rho_{air} V_{h|air}}}{\sum_{h' \in H} e^{\rho_{air} V_{h'|air}}}$	$IV_{air} = \log\left(\sum_{h' \in H} e^{\rho_{air} V_{h'|air}}\right)$
Air/non–air	$V_{air} = \omega X_{air} + \frac{1}{\rho_{air}} IV_{air}$ $V_{non-air} = 0$	$P_{air} = \dfrac{e^{V_{air}}}{1 + e^{V_{air}}}$	

The equations in Table 4.2 provide the conditional probabilities for the estimation of P_i. The saturated demand T_o (Equation 4.1) is modeled as the product of the population at the origin and a multiplicative factor, i.e., $T_o = Pop_o\alpha$, where α represents the maximum trips per capita that can be achieved by boosting connectivity and making transportation as seamless as possible. This coefficient (α) can be either empirically estimated as a function of socioeconomic features (e.g. GDP) or assumed to be equal to a given maximum trip potential. Following similar reasoning as in Hsiao and Hansen (2011), we assume α to be fixed and equal to 1 (i.e., one trip per capita per month), and test the consistency and robustness of the results through sensitivity analysis[6]. Branching down the tree, the saturated demand is split between air and non-air alternatives, with the latter representing the "outside good" for the choice situation that encompasses both the no-travel and travel-by-other-modes alternatives. Following the customary procedure (Berry 1994), the utility for the non-air alternative, $V_{non-air}$, is set to zero[7].

4.5   Model specification and estimation issues

4.5.1   Estimation

NL models can be estimated either simultaneously using maximum likelihood or sequentially by decomposing the NL tree into separate multinomial nested logit (MNL) models from the bottom to the top level. Although the simultaneous approach has proven to be more efficient than the sequential approach, the latter still provides consistent estimates and it has been largely used when dealing with aggregate data

---

[6]Hsiao and Hansen (2011) take the geometric mean of the populations in a city-pair market multiplied by a proportionality coefficient (arbitrarily set to 10 per quarter, i.e. about three potential trips per capita per month) to approximate the saturated demand in each city-pair market. Consistent with our origin-based framework, we consider the population at the origin instead. Although simple, this approach does not compromise the estimation of consistent parameters; for relatively high values of α, variations in the multiplier virtually only affect the intercept coefficients due to the difference-in-difference estimation resulting from the linearization of the MNL formula (refer to Hsiao (2008) for a more rigorous demonstration).

[7]The normalization of an outside good's utility is commonly reported in the literature as a method for overcoming a lack of information on unobserved alternatives and consistently estimating the model parameters.

(Allenby et al. 1991, Forinash and Koppelman 1993, Train 2003). In this paper, we rely on a sequential approach to enable the bootstrap procedure outlined in 4.5.3, which must be conducted separately for the bottom nests (itinerary choice).

The first step in the sequential approach is to estimate the lower level models (choice within a nest). The estimated coefficients, including the scale parameters, are then used to compute the inclusive values to be entered into the deterministic utility component of the parent nests in the upper model (choice between nests) (Train 2003, Hensher et al. 2005). By applying a linear transformation to the $P_{i|d,h,air}$ equation (Table 4.2), the vector of itinerary choice coefficients $\beta' = \mu\beta$ is obtained from the following equation (conditional subscripts omitted):

$$\log P_i - \log P_{i'} = \mu\beta(X_i - X_{i'}) + \xi_{ii'} \qquad \forall i, i' \in I_{od} \tag{4.4}$$

where $\xi_{ii'}$ is the difference between the unobserved utility components of itineraries i and i'. Treating $\xi_{ii'}$ as an unobserved error term, Equation 4.4 can be estimated by regressing the difference between the log market shares on the difference-in-attribute variables. Similarly, the $P_{d|h,air}$ equation (Table 4.2) can be transformed as follows:

$$\log P_d - \log P_{d'} = \lambda_h\gamma(X_d - X_{d'}) + \frac{\lambda_h}{\mu}(IV_d - IV_{d'}) + \xi_{dd'} \qquad \forall d, d' \in D_{oh} \tag{4.5}$$

The ratio of scale parameters is given by the coefficient of the difference between the inclusive values, which reflects the degree of intra-nest correlation and the extent to which changes in lower branches affect the choice probability of their nest.

Similar to $P_{i|d,h,air}$ (Equation 4.4) and $P_{d|h,air}$ (Equation 4.5), $P_{h|air}$ and P_air can also be linearized and sequentially estimated using linear regression models. The calibration dataset used in this study features a number of cross-sectional units, i.e., flight itineraries clustered by origin area, resulting in the likely violation of cross-sectional independence. Accordingly, we rely on Driscoll and Kraay's standard errors (Driscoll and Kraay 1998, Hoechle 2007), which are robust to heteroskedasticity as well as general forms of temporal and cross-sectional dependences.

4.5.2 Variable selection

As clarified in Section 4.4, under an NL formulation, specific explanatory variables can be entered at different stages of the tree structure. In this study, sets of variables are identified at the itinerary, destination, and origin levels, and a linear-in-parameters utility specification is used. The itinerary-specific variables (X_i) are the airfare, frequency, flying and connecting times, and type of service (non-stop vs one-stop). These factors were selected for their ability to capture important determinants of itinerary choice while being generally available at an aggregate level for large samples. According to the literature, airfare is an essential element of itinerary choice (e.g., Lurkin et al. 2017, Lhéritier et al. 2018) that generally represents the major out-of-pocket outlay incurred by passengers along their overall air trip, and its inclusion within the utility formulation allows for inference of the implied values of attributes as a key output of discrete choice models. Service frequency is used as a proxy for the allocated capacity on a city-pair market and also provides an indication of the flexibility of travel alternatives at passengers' disposal. Along with the airfare, service frequency is a major discretionary factor that an airline can leverage to effectively increase its market share (Wei and Hansen 2005). Type of service, flying time, and connecting time capture a number of multifaceted aspects of passenger utility when facing the trade-off between hasty-fast and comfortable-tardy connections. Additionally, we introduce airline dummies to represent preferences for a given airline due to relevant choice factors that are not made explicit in the itinerary utility function (e.g., frequent flyer programs, comfort, reliability, etc.) (Ishii et al. 2009).

At the destination level, demographic and socioeconomic destination-specific variables, namely the population of the metropolitan area to which access is granted and its economic activity measured in terms of the GDP per capita, are considered to reflect the destination quality (Allroggen et al. 2015). Further factors may be considered to capture, for example, the tourism orientation of each city or trade flows between origin and destination areas. As retrieving these variables for every metropolitan area is generally difficult, especially on a worldwide scale, destination fixed effects are

included in the model to account for unobserved features and specificities of destinations that affect their overall level of attractiveness. Supply-side variables, such as travel times and costs, are considered in the destination choice by the inclusive value computed based on the itinerary choice model.

Similarly, the air travel nest utility for each origin is defined as a function of the inclusive value derived from destination sub-nests, which combine the breadth of destinations available and the quality of air transport services to reach them, and origin dummies to capture intrinsic unobserved features that affect an origin area's ability to generate air trips (e.g., income, availability of other modes, etc.). This approach accommodates different substitution patterns between the air travel nest and the outside good, therefore obtaining origin-specific demand elasticities to changes in air transport supply.

4.5.3 Treatment of endogeneity and missing data

As noted in the previous section, airfare is widely acknowledged to be a major determinant of air travel behavior that should not be ignored in the modeling of air travel itinerary market share. From a practical point of view, the direct inclusion of airfare in the itinerary utility function poses two main challenges.

First, collecting proper and complete data reflecting the cost of flying is generally difficult. As a result of the complex and dynamic pricing strategies adopted by air carriers, it is both problematic and data-intensive to compute reliable average fares paid by customers based on posted prices. Flight data sources such as the Marketing Information Data Tapes (MIDT) provided by the OAG corporation contain consistent fare data at the itinerary-carrier level, although these are only partially complete. As missing airfare values are likely to be Missing at Not Random, e.g., most LCCs do not rely on the global distribution system (GDS), removing observations with missing values (listwise deletion) can produce a bias and yield unreliable parameter estimates (Rubin 1976, Saunders et al. 2006).

The second concern refers to the potential endogeneity between airfare and itinerary

choice[8]. (Mumbower et al. 2014, Lurkin et al. 2017). This is an important issue as it prevents us from directly entering observed airfare values into the OLS regression outlined in Equation 4.4.

To overcome these two problems, we propose a bootstrap procedure that combines regression-based imputation with the treatment of endogeneity. For N random samples drawn (with replacement) from the original data, the following three steps are performed on each bootstrap data set:

1. Estimate a first stage regression to proxy the airfare, using cases with complete data to calibrate the regression equation. Among the relevant predictors used—including the LCC and time dummies, which respectively capture low-cost business orientation and seasonality as important predictors of airfare—the product of distance and jet fuel cost is considered. According to Hsiao and Hansen (2011), this is a valid instrumental variable as it is highly correlated with airfare and not confounded with market share.

2. Use the estimated coefficients to perform imputation over all observations—both complete and non-complete cases. This produces an instrumented price that extrapolates the orthogonal component of price to the error term and is no longer endogenous to the market share.

3. Estimate the itinerary choice regression (Equation 4.4) replacing the observed price with the instrumented price as a predictor.

This process builds a sample of replicated beta estimates that constitute a sampling distribution of regression coefficients to be used for estimating standard errors and drawing statistical inference. The approach outlined in steps 1–3 above implies

---

[8]Other endogenous relationships may exist, specifically related to service attributes, such as flight frequency. However, in light of the dichotomy of supply and demand in air transport markets, such that the supply provided by each flight leg is shared by many city-pair markets (Belobaba 2009), and considering the exogenous constraints that to a large degree characterize frequency planning in practice (e.g. slot availability, mandatory flights, etc.), the problem of supply endogeneity appears limited and is tentatively not addressed in this study. Following Gayle (2013), we include airline dummies in the itinerary mean utility function, such that only non-airline-specific unobserved characteristics are captured by the error term, therefore making the assumption that observed service attributes are uncorrelated to unobserved quality even more reasonable.

imputation on each dataset, which, according to Shao and Sitter (1996), yields consistent bootstrap estimates and a better modeling of the uncertainty present in the missing values. Additionally, bootstrapping is an alternative approach to overcoming the main limitation of sequential 2SLS, i.e., the inability to obtain proper standard errors, when simultaneous 2SLS is not applicable (Guan 2003, Guevara-Cue 2010).

4.6   Empirical setting

4.6.1   Data collection

In this section, we describe the data used to assess the proposed NL model. A comprehensive data set for originating outbound trips from Italy covering a tenmonth period from January to October 2018 was assembled from four main data sources.

To evaluate the air trips generated from each metropolitan area and track their allocation over the full range of available air travel alternatives, data on O&D passenger flows and one-way airfares were collected from the OAG Traffic Analyser by selecting information for itineraries from any Italian airport to all final destinations worldwide and restricting the sample based on point of sales (POS) to originating outbound trips. For example, Figure 4-5 represents the portfolio of destinations that could be reached from Rome in January 2018. The most detailed information provided by the OAG Traffic Analyser is made available at the itinerary-carrier level on a monthly basis. For instance, the tuple (LIN, FCO, JFK, AZ, 201801) uniquely indexes the one-stop itinerary from Milan Linate (LIN) to New York John F. Kennedy (JFK) through Rome Fiumicino (FCO) operated by Alitalia (AZ) during January 2018[9].

Two filtering rules were applied to remove non-significant observations and ensure the reliability of the empirical analysis: only non-stop and one-stop itineraries were considered, which represented about 98.8% (83.6% non-stop and 15.1% one-stop) of total trips from Italy in 2018, and we only used itineraries that consistently carried

---

[9]Due to limitations arising from the use of aggregate data, a lack of customer information prevented us from segmenting them according to their purpose of travel.
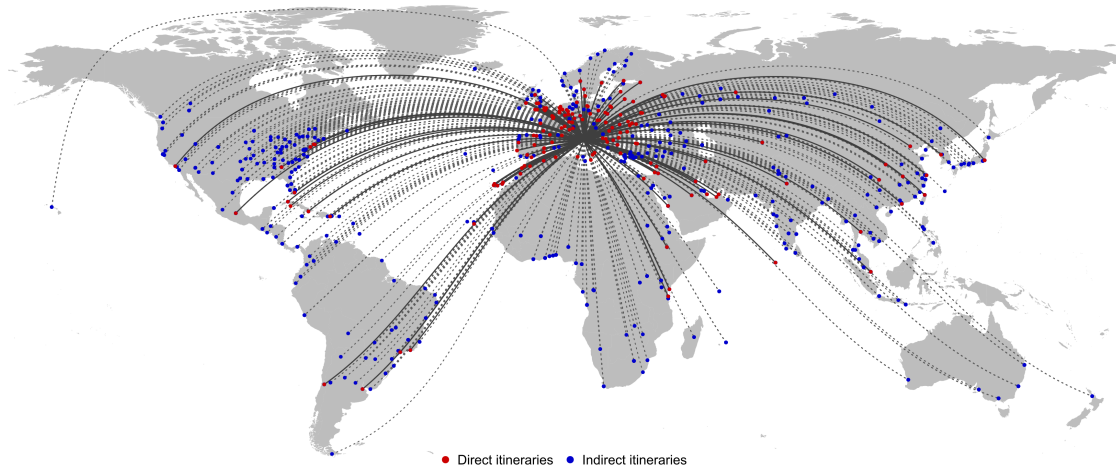
Figure 4-5: Reachable destinations from Rome (January 2018).

O&D traffic over at least six months to identify significant patterns of connectivity
and remove sporadic connections[10].

After selecting the relevant itineraries, each observation was complemented with
detailed data on transport supply (scheduled flight time, frequency, and connecting
time) sourced from the OAG Schedule Analyser. Although this information is readily
available for non-stop flights, it is not easy to obtain for connecting flights, for which
it was therefore calculated based on the schedules of the respective flight legs. In our
study, links were formed by relying on an itinerary building procedure that selected
the quickest path for any given scheduled departure time (e.g., Malighetti et al. 2008).
This approach accounts for the variety of alternatives at passengers' disposal without
overestimating the frequency of connecting itineraries. Additionally, a loose cut-off
condition was set to limit connecting times to within 1-6 h.

Demographic data were extrapolated from the latest version of the Gridded Pop-
ulation of the World collection (GPWv4), produced by the Socioeconomic Data and
Applications Center (SEDAC) at Columbia University, which contains disaggregated
population data at a resolution of 30 arc-seconds. For simplicity, we considered the
population of each metropolitan area as the number of people living within a fixed

---

[10]As many O&D markets, particularly those that are only served by indirect flights, are thin, we
opted for a filtering strategy based on recurrent connectivity patterns rather than setting an absolute
cut-off condition on the absolute volume of O&D air passenger flows.

radius of 50 km "as the crow flies". Data on gross domestic product (GDP) per capita at the country level were retrieved from the World Bank's World Development Indicators (WDI) database.

4.6.2 Descriptive statistics

After data filtering and cleaning, the resulting sample set consists of 101,397 monthly itinerary-carrier observations corresponding to 12,831 unique itineraries and 3,934 airport pairs. Airports are grouped within metropolitan areas (multi-airport regions) according to the classification provided by OAG to identify 29 origin and 602 destination areas, resulting in 2,876 unique city pairs. The data sample contains 200 unique carriers, including both point-to-point low-cost carriers (31) and hub-and-spoke legacy carriers (169).

The empirical distribution of origin-destination air passenger flows is highly skewed and varies widely within the data sample. The monthly average air passenger flow is 1,077 with a standard deviation of 3,277. The busiest route in the data sample, Milan to London, features an average of 55,300 originating outbound passengers per month. On average, each originating area is connected to 117 destinations. The Italian metropolitan region with the highest number of destinations is Rome, which is included in 516 city pairs, of which 57 are connected via non-stop flights, 342 via one-stop flights, and 117 via both non-stop and one-stop flights (Figure 4-5).

In accordance with the objectives of the study, we divide the destinations into two length-of-haul categories: Medium-Haul (MH: 500 km $\leq$ distance $<$ 3,000 km), and Long-Haul (LH: distance $>$ 3,000 km). Very short-haul destinations (distances less than 500 km) are not considered in the analysis. Within the short haul travel range, air transport is subject to greater competition from other modes and to a higher degree of modal substitution (Lieshout et al. 2016), such that considering the outside good as a market fixed effect would be rather limiting. The empirical distribution of destinations served from Italy is represented as a function of distance (great circle method) in Figure 4-6. The distance threshold is set to 3,000 km to delineate the subsets of continental and intercontinental destinations, and a sensitivity analysis is

148

conducted to assess the robustness of estimates to different-cut-off values (See Section 7.2). Based on this partition, out of the 101,397 valid observations, 53,823 and 47,574 are designated as MH and LH, respectively.
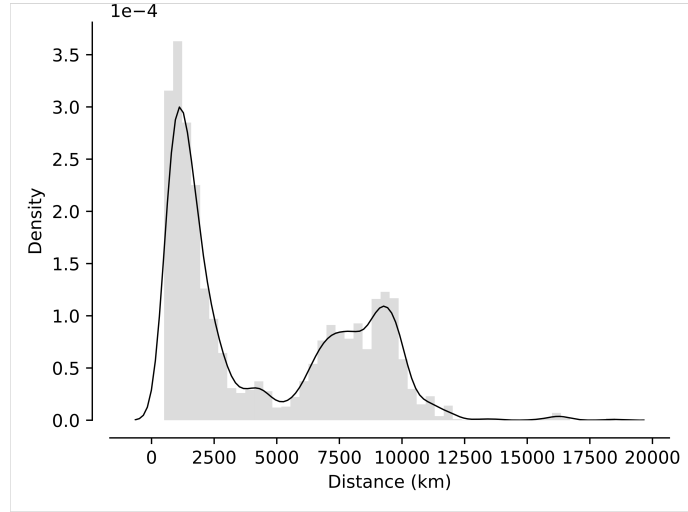

Figure 4-6: Distribution of destinations as a function of distance (km)

Table 4.3 lists descriptive statistics related to the itinerary choice variables. In general, an average non-stop itinerary occupies more than the double the market share of a one-stop route (57% vs 22%) but with a much lower frequency (-37% on LH vs -44% on MH). However, the role of connecting itineraries cannot be neglected, as demonstrated by their relevant contribution to both the LH and MH markets. One-way airfares have reasonable values ($560 and $123 for an average one-stop flight on LH and MH routes, respectively) and exhibit high variability, particularly among connecting itineraries. Similar to other studies in the literature (Mumbower et al. 2014), a major problem with our dataset is the significant amount of missing airfare data (31,398 records, i.e., 30.9%). As detailed in Section 4.5.3, we implement an ad-hoc procedure to address this problem and ensure the reliability of regression estimates.

Table 4.3: Descriptive Statistics – Itinerary choice.

Haul-length	Itinerary	Variable	Mean	Std dev	5%	50%	95%
MH	Non-stop	Market share	57.3%	31.6%	9.3%	55.1%	100.0%
		Flying time	130.2	41.4	80	120.4	210
		Connecting time	-	-	-	-	-
		Frequency (monthly)	29.6	35.3	4	17	101.3
		Price	122.6	59.2	47	113	233.9
	One-stop	Market share	22.2%	28.6%	0.3%	8.9%	99.2%
		Flying time	213.4	54.5	135	210	315
		Connecting time	156.2	59.2	74.3	150	261
		Frequency (monthly)	52.4	48.1	4	37.2	136
		Price	169.3	111.9	56	144	370.3
LH	Non-stop	Market share	57.1%	26.8%	12.8%	56.4%	100.0%
		Flying time	528.6	168	275	565	755
		Connecting time	-	-	-	-	-
		Frequency (monthly)	18.7	16.2	3	14	40.9
		Price	489.2	241.2	150.1	476	915.8
	One-stop	Market share	21.5%	24.9%	1.0%	11.9%	88.0%
		Flying time	712.1	156.4	408.4	730	936
		Connecting time	170.9	64.7	78.9	165	290
		Frequency (monthly)	29.7	24.2	4.4	25.3	74.2
		Price	560.2	535	114	418	1482.1

4.7 Results

4.7.1 Econometric results

Table 4.4 reports the results of four different models used to assess the validity of our modeling approach. Model 1 (NL-ALL OLS) estimates all stages sequentially using OLS without differentiating by length of haul. Model 2 (NL-ALL) applies the bootstrap procedure to examine the effects of adjusting for endogeneity and missing data. Models 3 (NL2-MH) and 4 (NL2-LH) report the results of the NL2 formulation for medium- and long-haul trips, respectively. As they were not generally consistent with utility maximization theory, the results of NL1 are not reported; by contrast, for the sample at hand the NL2 formulation provides robust and consistent estimates, i.e., log-sum coefficients between zero and one, suggesting that air trip generation is rather independent across lengths of haul and, therefore, that separate modeling is preferable. The saturation demand multiplier (α) and the number of bootstrap samples are arbitrarily set to 1 and 500, respectively (see Section 4.7.2 for robustness

tests).

Table 4.4: Estimation Results.

Nest Level	Variable	(1)NL-ALL *OLS*	(2)NL-ALL *BTSRP-IV*	(3)NL2-MH *BTSRP-IV*	(4)NL2-LH *BTSRP-IV*
Itinerary	Frequency (flights per month)	0.0118*** [0.0000]	0.0110*** [0.0003]	0.0104*** [0.0003]	0.0111*** [0.0007]
	Flight time (minutes)	-0.0056*** [0.0000]	-0.0054*** [0.0002]	-0.0113*** [0.0005]	-0.0040*** [0.0002]
	Connecting time (minutes)	-0.0032*** [0.0000]	-0.0036*** [0.0001]	-0.0043*** [0.0002]	-0.0030*** [0.0001]
	Type of service (=1, non-stop)	2.6203*** [0.0610]	2.5921*** [0.0320]	2.2320*** [0.0473]	2.4272*** [0.0640]
	Price (2018 US$)	7.60E-05 [4.87E-05]			
	Price (IV) (2018 US$)		-0.0159*** [0.0017]	-0.0171*** [0.0019]	-0.0012 [0.0009]
	Airline fixed effects	YES	YES	YES	YES
	R2 adjusted	0.574	0.521	0.658	0.3028
Destinations	Population (millions of people)	0.0407*** [0.0048]	0.1009*** [0.0038]	0.0940*** [0.0032]	0.0272*** [0.0032]
	GDP per capita ('000 US$)	-0.0227*** [0.0021]	-0.0043* [0.0025]	0.0065*** [0.0012]	0.0124*** [0.0020]
	Inclusive value ($\frac{\rho_{air}}{\mu}$)	0.9858*** [0.0280]	0.7736*** [0.0250]	0.8257*** [0.0219]	0.8995*** [0.0230]
	Destination fixed effects	YES	YES	YES	YES
	R2 adjusted	0.877	0.826	0.851	0.767
Air/non-air	Constant	-6.2041*** [0.1800]	-5.2511*** [0.2816]	-5.0894*** [0.1880]	-7.8968*** [0.1165]
	Inclusive value ($\frac{1}{\rho_{air}}$)	0.8189*** [0.0430]	0.6971*** [0.0572]	0.8307*** [0.1119]	0.7516*** [0.0698]
	Origin fixed effects	YES	YES	YES	YES
	R2 adjusted	0.979	0.959	0.964	0.93

Confidence levels: *$p < 0.10$. ** $p < 0.5$. *** $p < 0.01$.
Note: Driscoll and Kraay' standard errors into brackets.

As shown in Table 4.4, the use of OLS in Model 1 yields an insignificant coefficient for airfare as a bias that arises when endogeneity is ignored. The bootstrap procedure, on the other hand, appears to properly address the issue of confounding and missing data to produce more reasonable results. A comparison of the results of Model 2 with those of Models 3 and 4 reveals that consideration of trip length highlights significant differences in sensitivity to travel determinants and, accordingly, helps clarify trip generation and allocation dynamics.

Overall, the model coefficients are statistically significant and their signs are consistent with expectations. Both the frequency and type of service have a positive

impact on itinerary market share and neither varies notably across trip length (Figure 4-7). Increasing either the connecting or flight times worsens the attractiveness of itineraries. In absolute terms, their impacts are found to be more burdensome at shorter distances; specifically, the marginal effects of connection and flight times on LH flights are estimated to be 30 and 65% less, respectively, than on MH flights (Figure 4-7). In addition to having an overall longer trip duration, LH demand is heavily reliant on connecting traffic, with 63% of LH traffic involving a connection (7% domestically) against 37% flying non-stop; the corresponding split for MH flights is 7% (40% domestically) and 93%, respectively. Given the scattered distribution of LH point-to-point markets, passengers on LH travel are in general more willing to accept longer travel times and to consider indirect itineraries as a valid travel option.
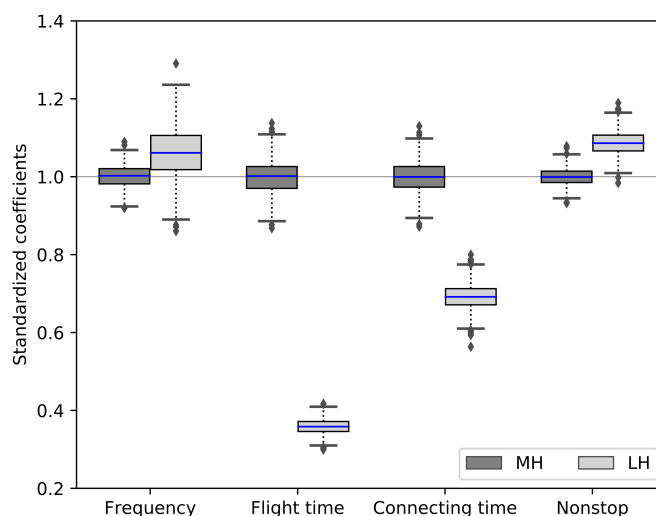

Figure 4-7: Comparison of coefficients – MH vs LH (normalized to MH coefficients).

For MH flights, airfare is found to have a significant negative impact on the itinerary choice; although the impact for LH flights is also negative, it appears to be negligible. This can be explained by intrinsic features of LH travel and the data samples used. Price elasticity on LH flights is expected to be lower as a result of the shortage of inter-modal alternatives and the higher typical cost of intercontinental transportation (for a detailed overview of the determinants of price elasticities, see Brons et al. (2002)). Additionally, LH flight choice is likely to be more affected by

the business component in terms of both price differential over the leisure counterpart and a reduced propensity to switch to economy class "no-frills" service (Francis et al. 2007). With regards to the empirical setting, the LH segment is highly heterogeneous, encompassing geographic markets with both low and high elastic price responses (e.g., Europe–Asia and Europe–North America, respectively) (Smyth and Pearce 2008). Intra-Europe routes, by contrast, are characterized by a high degree of substitution, strong competition, and price-sensitive demand.

Model 3 produces a consistent adjusted R^2 value of 0.66, demonstrating that the included explanatory variables are able to accurately capture itinerary choice behavior in MH markets. The explanatory variables are also relevant for LH flights, although their compound explanatory power is smaller (0.30), reflecting the important role played by additional features and the fact that more frills are required in the LH sector (Francis et al. 2007).

To facilitate a comparison of coefficients in terms of magnitude, we compute the value of time (VOT) measures under those models reporting significant betas for airfare. For Model 2 (NL-ALL), the average VOTs are 20.4 and \$13.6/h for flight and connecting times, respectively. These are reasonable values and of the same order of magnitude as those reported in the literature[11]. In the MH markets, these VOTs increase to 39.7 and \$15.1/h, respectively. In terms of frequency, an additional daily flight is assigned a marginal equivalent value of \$18.3/h, while the inferred monetary value for non-stop itineraries is \$130.5/h.

At the destination level, the GDP per capita and population are found to positively impact the attractiveness of destinations, with the model indicating a good fit in terms of adjusted R^2—0.85 and 0.77 for NL2-MH and NL2-LH, respectively—as a result of the fixed effects in the model. The nesting coefficients are considerably high, particularly for the LH markets, demonstrating the significance of the origin-based approach. Although itineraries serving the same city pair are correlated—with

---

[11]For instance, Hsiao and Hansen (2011) report different VOTs for scheduled flight times on direct and indirect routes of 16.8 and \$24.1/h, respectively, while Lurkin et al. (2017) compute the value of time for the total elapsed time for high- and low-yield products, equal to 83.30 ad \$43.36/h, respectively.

nesting coefficients below one, confirming the validity of the nesting structure—their impact on destination choice is not negligible. Variations in itineraries such as the introduction of a new flight, lower prices, and improved service levels do not solely shift traffic between competing itineraries but considerably affect passenger distribution among destinations.

In the trip generation stage, the scale parameters, i.e., $\frac{1}{\rho_{air}}$, are statistically significant, consistent with utility-maximization theory, and moderate in magnitude (0.83 and 0.75 for MH and LH, respectively). Thus, the introduction of new destinations and the increased number of flight options in certain markets alter the destination market shares while also significantly impacting the total number of people traveling by air.

4.7.2 Robustness checks

This section presents three robustness checks to corroborate the results of the empirical analysis. In detail, we test how changes in key assumptions of the model, i.e., the length of haul threshold, the saturated demand multiplier (α), and the number of bootstrap samples (N), affect the NL2-estimated coefficients presented in Table 4.4. Table 4.5 reports the itinerary choice coefficients for different MH/LH threshold values (3,000 $\pm$ 500 km). The stability of these results confirms the validity of the length-of-haul partition for the Italian case.

Table 4.5: Sensitivity Analysis – MH/LH distance thresholds

Haul-length		Frequency		Flight Time		Connecting Time		Type of service		Price (IV)	
		coef	*std*	*coef*	*std*	*coef*	*std*	*coef*	*std*	*coef*	*std*
MH	$\leq 2,500$	0.011	0.000	-0.011	0.001	-0.004	0.000	2.252	0.054	-0.015	0.001
	$\leq 3,000$	0.010	0.000	-0.011	0.001	-0.004	0.000	2.232	0.047	-0.017	0.002
	$\leq 3,500$	0.011	0.000	-0.011	0.000	-0.004	0.000	2.222	0.040	-0.017	0.002
LH	$> 2,500$	0.011	0.001	-0.004	0.000	-0.003	0.000	2.423	0.064	-0.004	0.001
	$> 3,000$	0.011	0.001	-0.004	0.000	-0.003	0.000	2.427	0.064	-0.001	0.001
	$> 3,500$	0.011	0.001	-0.004	0.000	-0.003	0.000	2.461	0.071	-0.001	0.001

Another relevant assumption to investigate is the maximum number of monthly trips per capita (α). As shown in Table 4.6, when α is varied from 1 (one trip per

person per month) to 5 (five trips per person per month), the air travel stimulation coefficients, i.e., $1/\rho_{air}$, remain statistically significant and their magnitude is largely unaffected, while the constant term changes widely. Given our goal of demonstrating the significance of adopting an origin-based approach to demand forecasting, this corroborates the findings discussed in the previous section. However, studies with different purposes may require other approaches to evaluate α and each region's saturation demand.

Table 4.6: Sensitivity Analysis – Saturation demand parameter (α).

α	MH				LH			
	Inclusive value		Constant		Inclusive value		Constant	
	coef	std	coef	std	coef	std	coef	std
1	0.8307	0.1119	-5.0894	0.188	0.7516	0.0699	-7.8968	0.1165
2	0.7873	0.1034	-5.706	0.1691	0.7514	0.0698	-8.5903	0.1165
3	0.7749	0.1011	-6.0903	0.1641	0.7513	0.0698	-8.9958	0.1165
4	0.7689	0.1000	-6.3681	0.1619	0.7513	0.0698	-9.2835	0.1165
5	0.7654	0.0994	-6.5855	0.1605	0.7513	0.0698	-9.5067	0.1165

Figure 4-8 illustrates the coefficients' sampling distribution for different numbers of bootstrap samples (100, 500, and 1,000). While the green distributions (100) display irregular patterns and produce unstable coefficient estimates, the red (500) and blue (1,000) lines are "normal-shaped" and nearly identical. Ultimately, we selected a number of random draws equal to 500 as this provides sufficiently accurate (and stable) distributions while reducing the computation time[12].

4.8   Conclusions

Over the past few decades, air travel estimation has received a good deal of interest from both academics and practitioners developing tools and quantitative methods to address the challenges and the increasing complexity of the aviation industry. This paper makes three main contributions to the modeling of air travel demand: (1) it proposes a new integrated origin-based air travel demand model that considers the

---

[12]The iterative bootstrap procedure outlined in Section 4.5.3 was implemented in Python 3.6, using the Statsmodels package [www.statsmodels.org] for econometric implementation.
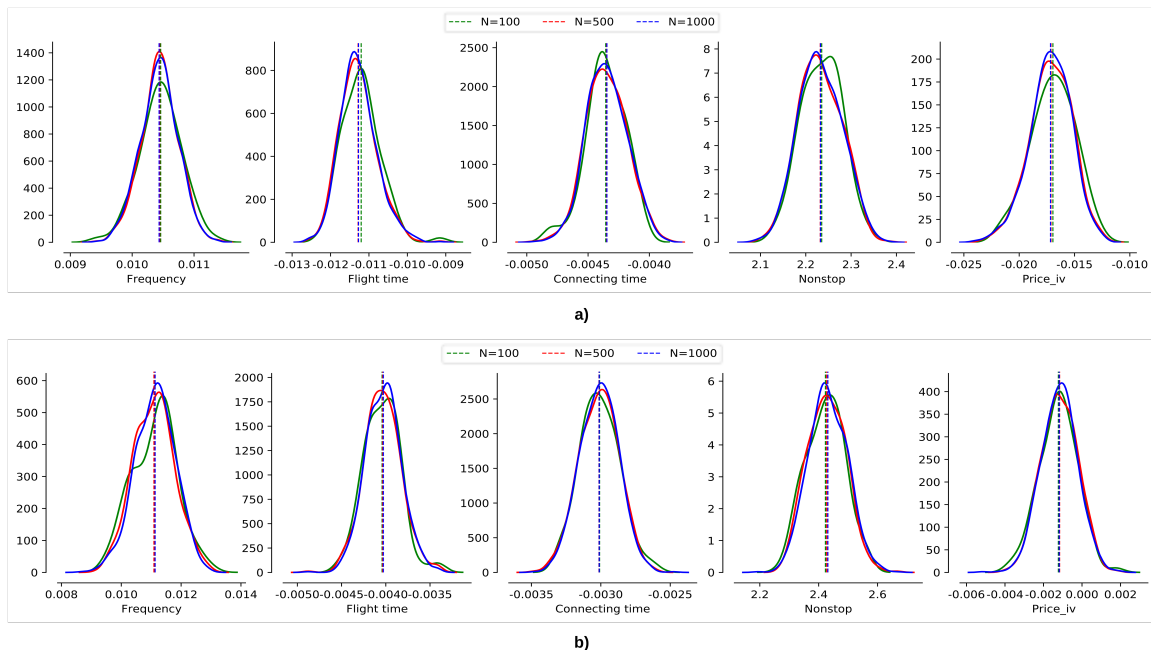
Figure 4-8: Sensitivity analysis – Number of bootstrap samples. (a). Medium-haul, (b). Long-haul.

originating travel forces of each area and explicitly considers substitutability between alternative destinations; (2) it develops two general aggregate NL formulations to reflect the level of independence between lengths of haul and account for differences in passengers' choice behavior and sensitivity to itinerary attributes; (3) it applies an ad-hoc bootstrap procedure to jointly address endogeneity arising between airfare and itinerary market share and data missingness.

The proposed integrated model is tested over the entire network of outbound air trips from Italian airports in 2018, with the findings indicating that medium- and long-haul trips exhibit considerably different features. Connecting time, flight time, and airfare play a more decisive role in itinerary choice for medium- than for long-haul trips, while frequency and type of service are found to have similar impacts on the two haul distances. Air transport supply is found to significantly affect the destination choice and potentially stimulate new air traffic.

The implementation of the proposed integrated demand model can support actors in the industry in carrying out planning activities and strategic decision-making. Airports might benefit from a thoughtful assessment of outbound trip generation and

demand saturation that would enable them to detect the most promising trajectories for development. Airlines could leverage an origin-based approach to obtain valuable insights into route development and network planning. Policymakers could benefit from the estimation of newly generated demand and passenger distributions among both itineraries and destinations to better assess the consequences of stimulation (or restriction) policies aimed at enabling (or hindering) the expansion of the aviation sector.

However, the proposed model is not without limitations. One major limitation follows directly from the lack of information on other modes of transport, which kept us from assessing alternative nesting structures. In particular, the grouping of alternatives first by destination and then by air/non-air routes was not possible because of a lack of specific information on the relevance of intermodal competition in each O&D market.

Promising avenues for future research lay in increasing the model's estimation accuracy and applicability. First, one approach will be to segment air trips into homogeneous groups according to their travel purpose, e.g., leisure vs business, and journey type, each of which are important features affecting the travel decision-making process. Second, a promising research direction will involve further development of the saturated demand function. Although not strictly necessary in the investigation of trade-offs between choice determinants and substitution patterns within a multi-level NL framework, the careful assessment of each region's trip generation potential is key to inferring stimulated trips and producing reasonable traffic estimates when substantial variations are made to air transport supply. Third, the inclusion of airport ground access (and egress) considerations will further improve the model by accounting for passenger allocation over neighboring airports within the same metropolitan area and tracking each step along the comprehensive air passenger trip from their originating point to their final destination. Fourth, given that this research focuses on the demand function and does not jointly estimate a supply function, the more explicit treatment of supply and quality choices can be subject of future research.

Chapter 5

Conclusion

This thesis has addressed the modeling of supply-demand interactions in the optimization of air transport networks. Starting from the longest-term strategic decisions to daily schedules, demand represents perhaps the most important input to the design, planning, and allocation of aviation resources. Demand, however, is hard to predict and it is inextricably tied to air transport supply. While the added value of capturing the two-way relationship between supply and demand is not under debate, the explicit modeling of supply-demand interactions requires addressing a number of challenges—both related to the empirical estimation of advanced demand models as well as their integration into optimization models supporting the airline planning process—and has been substantially underrepresented in the literature. This thesis aims to contribute to this important stream of research by developing integrated models that organically integrate demand modeling and mathematical programming. Throughout Chapters 2-4, we have illustrated these models and highlighted their capability to deliver valuable insights and practically aid decision-making.

The first contribution (Chapter 2) has dealt with airline network planning. In this paper, we have developed an original modeling framework that captures the interdependencies between supply (service frequency and network structure) and demand—referred to as Airline Network Planning with Supply and Demand interactions (ANPSD). The ANPSD contributes to the literature in three major ways. First, decision support tools are much less prevalent in airline strategic planning, as compared

to the pervasive use of operations research in support of other stages of the airline planning process. In this respect, the ANPSD constitutes probably the first data-driven approach that comprehensively and simultaneously addresses network planning problem decisions, spanning route planning, flight frequency and fleet composition, capturing the interdependencies between demand and supply in airline strategic planning. Second, the demand model has been estimated empirically based on historical data and the form of the empirical demand function has led the ANPSD to be formulated as a non-linear, non-convex mixed-integer optimization problem—which is highly challenging to solve. Hence, we have developed an original exact cutting-plane algorithm, named 2αECP, which leverages the structure of the demand function but also represents a novel solution method for a broader class of nonconvex MINLPs. The validity of the proposed approach has been evaluated through a set of comprehensive computational experiments based on real-world-like instances and a case-study involving the continental network of Alitalia. Computational results suggest that the 2αECP algorithm yields stronger solutions than state-of-the-art benchmarks based on discretization and linearization, and, after two hours of computation time, terminates within 1%–5% optimality gaps in realistic instances. Finally, practical results demonstrate that the ANPSD enacts more rationale and meaningful strategic insights, as compared to conventional approaches relying on inelastic demand, or entailing only a partial representation of supply-demand interactions. Results also show how ANPSD practically aids decision-making concerning network expansions, being able to suitably appraise the most promising development opportunities and drive the synergic evolution of fleet and network.

The second contribution (Chapter 3) has dealt with airline tactical planning. In this paper, we have developed a novel mixed-integer nonlinear flight scheduling and fleet assignment optimization model wherein air travel demand generation and allocation are simultaneously and consistently endogenized (IGASFAM). Different from the ANPSD, the optimization literature on airline tactical planning is rather vast and the study of supply-demand interactions has been subject to a number of (recent) studies. Yet, these contributions have focused primarily on the demand allocation problem,

using basic "S-curve" or MNL models to express the relationship between supply and demand. As a main drawback, these models do not properly capture (or do not capture at all) the demand generation (or degeneration) effects of increased (or decreased) level of service. The IGASFAM significantly contributes to this body of knowledge by enhancing the representation of supply-demand interactions to simultaneously consider both demand generation and allocation dynamics. This is achieved by using a nested logit models that simultaneously capture demand generation and allocation among air travel itineraries. To deal with the resulting nonlinearities in the mathematical formulation, a piecewise linearization scheme and tightening constraints are introduced, which improve the tractability of the model and make it solvable for rather large instances. The validity of the proposed model has been demonstrated by both computational experiments and a real-world case study involving a major European airline (Alitalia). Results have demonstrated how the IGASFAM can solve mid-size hub-and-spoke networks in reasonable times and that substantial profit gains (+6.9%) could be achieved through its application, thus representing a valuable tool for supporting the airline planning process.

Finally, in the third contribution (Chapter 4), we have proposed a novel demand model for air transportation. This paper contributes to the literature by addressing a key limitation of existing modeling approaches, which estimate air travel demand independently within city-pairs. This formulation is quire restrictive in that it assumes that air travel between two cities is not affected by changes in air transport supply elsewhere. However, if the two markets are overlapped (i.e. they share the same origin or destination) it is likely the case that adding or removing destinations may affect the resulting flows in the other markets. More importantly, taking a city-pair perspective overlooks the saturation of demand at the origin, resulting in the over- or under- overestimation of the number of trips generated from a given region when substantial variations are made to air transport supply (e.g., the addition or deletion of routes from a given origin area). To address this issue, this paper has developed an origin-based air travel demand model that assumes saturation at the origin level and explicitly accounts for substitutability between destinations. The proposed approach

simultaneously integrates demand generation and allocation using a multilevel aggregate nested logit formulation that covers the choices of whether or not to travel by air, where to travel (destination), and how to travel (itinerary). Two specifications have been proposed to reflect systematic differences between lengths of haul and the bootstrap has been applied to jointly address endogeneity issues and data missingness. The output of the proposed model can serve as a useful planning and strategy tool by airports and airlines to detect the most promising development trajectories in both saturated and developing catchment areas. Also, policymakers could benefit from the estimation of newly generated demand and passengers diverting among both itineraries and destinations enabled by the model, for better assessing the welfare implications of stimulation (restriction) policies aimed at enabling (hindering) the expansion of the aviation sector.

Collectively, these three works contribute to the literature on air transport planning by developing tools and methodologies that are applicable to real-world cases and practically support decision-making through an advanced representation of the interdependencies between supply and demand. Furthermore, the modeling and algorithmic frameworks proposed in this thesis provide valid foundations for future research toward a systematic and comprehensive integration of supply-demand interactions in decision support tools for air transport network planning.

Besides the specific points identified in each chapter, three main directions for future research are identified.

First, the models presented in this thesis have all relied on parametric techniques to estimate demand models. Recently, the use of machine learning methods has also gained popularity in the OR literature (e.g., Jaquillat 2020, Beulen et al. 2020). Different from parametric models, machine learning methods do not provide a closed-form expression that can be directly embedded into mathematical formulations and do not inform on causal relationships. On the other hand, machine learning methods are acknowledged to better capture nonlinear and complex relationships among variables, potentially yielding superior prediction performances. The application of these methods in the context of supply-demand interactions is scant, but certainly

constitutes an interesting direction for future research.

Second, all the proposed approaches have considered a deterministic demand model. However, besides the simultaneity between supply and demand, another key aspect relates to the intrinsic uncertainty of demand, especially in the long term. Stochastic and robust optimization approaches have been subject to extensive research—also applied to airline planning (e.g., Marla et al. 2018, Cadarso and de Celis 2017). Yet, none of these contributions has developed a comprehensive modeling framework that incorporates both demand uncertainty and supply-demand interactions. This will requite to identify and characterize the determinants of air travel demand, and model uncertainties regarding their future evolution. Although challenging, both empirically and computationally, this approach could lead to a more effective mapping of uncertainty, ultimately resulting in superior decision making.

Third, an interesting research direction relates to the third contribution. The origin-based model presented in chapter 4 has been illustrated as a stand-alone demand model. Future works may be devoted, first, to evaluate the potential benefits that can be derived from its application in the context of network planning. Additionally, the ability to capture saturation at the origin level and correlations between destinations and itineraries, makes the origin-based model developed in this thesis particularly suitable for airport network planning. Airports in general have been much less active in route development than airlines. However, in the last years - and before the brutal halt due to Covid-19 - airports have been facing growing and pressing environmental concerns, and the scarcity of airport resources (compared to the exponential growth of air travel) has led to increasing congestion issues, especially at major hubs. Hence, it is always more crucial for airport planners and strategists to carefully plan the most promising directions for future developments in order to balance off the often conflicting interests related to the consolidation of hub operations and the overall welfare impact for the local community.

Air travel will recover from Covid-19 and hopefully restore to a sustained growth pattern. By developing advanced methodologies to account for supply-demand interactions, this thesis strives to contribute to the rebuilding of global air transport

networks and to the upcoming challenges facing the aviation industry.

Bibliography

Abara J (1989a) Applying Integer Linear Programming to the Fleet Assignment Problem. *Interfaces* 19(4):20–28.

Abara J (1989b) Applying Integer Linear Programming to the Fleet Assignment Problem. *Interfaces* 19(4):20–28.

Abed SY, Ba-Fail AO, Jasimuddin SM (2001) An econometric analysis of international air travel demand in Saudi Arabia. *Journal of Air Transport Management* 7(3):143–148.

Abrahams M (1983) A service quality model of air travel demand: An empirical study. *Transportation Research Part A: General* 17(5):385–393.

Adler N, Hashai N (2005) Effect of open skies in the Middle East region. *Transportation Research Part A: Policy and Practice* 39(10):878–894.

Adler N, Njoya ET, Volta N (2018) The multi-airline p-hub median problem applied to the African aviation market. *Transportation Research Part A: Policy and Practice* 107:187–202.

Adler N, Pels E, Nash C (2010) High-speed rail and air transport competition: Game engineering as tool for cost-benefit analysis. *Transportation Research Part B: Methodological* 44(7):812–833.

Allenby GM, Rossi PE, Allenby GM, Rossi PE (1991) There Is No Aggregation Bias : Why Macro Logit Models Work. *Journal of Business & Economic Statistics* 9(1):1–14.

Allroggen F, Wittman MD, Malina R (2015) How air transport connects the world – A new metric of air connectivity and its evolution between 1990 and 2012. *Transportation Research Part E: Logistics and Transportation Review* 80:184–201.

Andersson H, Hoff A, Christiansen M, Hasle G, Løkketangen A (2010) Industrial aspects and literature survey: Combined inventory management and routing. *Computers and Operations Research* 37(9):1515–1536.

Antunes A, Martini G, Porta F, Scotti D (2020) Air connectivity and spatial effects: regional differences in Europe. *Regional Studies* 0(0):1–13.

ATAG (2018) Aviation Benefits Beyond Borders. Technical report.

Atasoy B, Salani M, 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.

Barnhart C, Belobaba P, Odoni AR (2003a) Applications of operations research in the air transport industry. *Transportation Science* 37(4):368–391.

Barnhart C, Boland NL, Clarke LW, Johnson EL, Nemhauser GL, Shenoi RG (1998a) Flight string models for aircraft fleeting and routing. *Transportation Science* 32(3):208–220.

Barnhart C, Cohn A (2004) Airline schedule planning: Accomplishments and opportunities. *Manufacturing and Service Operations Management* 6(1):3–22.

Barnhart C, Cohn AM, Johnson EL, Klabjan D, Nemhauser GL, Vance PH (2003b) *Airline Crew Scheduling*, 517–560 (Boston, MA: Springer US).

Barnhart C, Farahat A, Lohatepanont M (2009) Airline Fleet Assignment with Enhanced Revenue Modeling. *Operations Research* 57:231–244.

Barnhart C, Johnson EL, Nemhauser GL, Savelsbergh MW, Vance PH (1998b) Branch-and-price: Column generation for solving huge integer programs. *Operations Research* 46(3):316–329.

Barnhart C, Kniker TS, Lohatepanont M (2002) Itinerary-Based Airline Fleet Assignment. *Transportation Science* 36(2):199–217.

Barnhart C, Lu F, Shenoi R (1998c) Integrated Airline Schedule Planning. 384–403 (Springer, Boston, MA).

Başar G, Bhat C (2004) A parameterized consideration set model for airport choice: an application to the San Francisco Bay Area. *Transportation Research Part B: Methodological* 38(10):889–904.

Belobaba P (2009) Overview of airline economics, markets and demand. Belobaba P, Odoni A, Barnhart C, eds., *The Global Airline Industry*, 47–72 (John Wiley & Sons, L td.).

Belobaba P, Odoni AR, Barnhart C (2015) *The global airline industry*.

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

Berry S, Jia P (2010) Tracing the Woes: An Empirical Analysis of the Airline Industry. *American Economic Journal: Microeconomics* 2(3):1–43.

Berry S, Reiss P (2007) *Empirical Models of Entry and Market Structure*, volume 3 (Elsevier).

Berry ST (1994) Estimating Discrete-Choice Models of Product Differentiation. *The RAND Journal of Economics* 25(2):242.

Beulen M, Scherp L, Santos BF (2020) Dynamic evaluation of airline Crew's flight requests using a neural network. *EURO Journal on Transportation and Logistics* (June):100018.

Bhadra D (2003) Demand for air travel in the United States: Bottom-up econometric estimation and implications for forecasts by origin-destination pairs. *Journal of Air Transportation* 8 (2)(2003):19–56.

Birolini S, Cattaneo M, Malighetti P, Morlotti C (2020) Integrated origin-based demand modeling for air transportation. *Transportation Research Part E: Logistics and Transportation Review* 142(June):102050.

Birolini S, Malighetti P, Redondi R, Deforza P (2019) Access mode choice to low-cost airports: Evaluation of new direct rail services at Milan-Bergamo airport. *Transport Policy* 73:113–124.

Boonekamp T, Zuidberg J, Burghouwt G (2018) Determinants of air travel demand: The role of low-cost carriers, ethnic links and aviation-dependent employment. *Transportation Research Part A: Policy and Practice* 112:18–28.

Brons M, Pels E, Nijkamp P, Rietveld P (2002) Price Elasticities of Demand for Passenger Air TravelRi, Hsduwphq. *Journal of Air Transport* 8(3):165–175.

Brueckner JK, Dyer NJ, Spiller PT (1992) Fare Determination in Airline Hub-and-Spoke Networks. *The RAND Journal of Economics* 23(3):309.

Brueckner JK, Lee D, Singer E (2014) City-Pairs Versus Airport-Pairs: A Market-Definition Methodology for the Airline Industry. *Review of Industrial Organization* 44(1):1–25.

Brueckner JK, Lee D, Singer ES (2013) Airline competition and domestic US airfares: A comprehensive reappraisal. *Economics of Transportation* 2(1):1–17.

Bryan DL, O'Kelly ME (1999) Hub-and-spoke networks in air transportation: An analytical review. *Journal of Regional Science* 39(2):275–295.

Burghouwt G, Redondi R (2013) Connectivity in Air Transport Networks An Assessment of Models and Applications. *Journal of Transport Economics and Policy* 47(1):35–53.

Business Insider (2020) Airlines face a years-long challenge to rebuild global route maps and networks that were devastated by the COVID-19 pandemic. https://www.businessinsider.com/airlines-and-air-travel-after-covid-19-coronavirus-pandemic-2020-6.

Button K, Drexler J (2005) Recovering costs by increasing market share: An empirical critique of the s-curve. *Journal of Transport Economics and Policy* 39(3):391–404.

Cacchiani V, Salazar-González JJ (2017) Optimal solutions to a real-world integrated airline scheduling problem. *Transportation Science* 51(1):250–268.

Cadarso L, de Celis R (2017) Integrated airline planning: Robust update of scheduling and fleet balancing under demand uncertainty. *Transportation Research Part C: Emerging Technologies* 81:227–245.

Cadarso L, Vaze V, Barnhart C, Marín Á (2017) Integrated airline scheduling: Considering competition effects and the entry of the high speed rail. *Transportation Science* 51(1):132–154.

Campbell JF, O'Kelly ME (2012) Twenty-Five Years of Hub Location Research. *Transportation Science* 46(2):153–169.

Carrasco JA, de Dios Ortúzar J (2002) Review and assessment of the nested logit model. *Transport Reviews* 22(2):197–218.

Carreira JS, Lulli G, Antunes AP (2017) The airline long-haul fleet planning problem: The case of TAP service to/from Brazil. *European Journal of Operational Research* 263(2):639–651.

Cattaneo M, Malighetti P, Redondi R, Salanti A (2018) Changes in frequencies and price variations on point-to-point routes: The case of easyJet. *Transportation Research Part A: Policy and Practice* 112:60–70.

Cho W, Windle RJ, Dresner ME (2015) The impact of low-cost carriers on airport choice in the US: A case study of the Washington-Baltimore region. *Transportation Research Part E: Logistics and Transportation Review* 81:141–157.

Chung JY, Whang T (2011) The impact of low cost carriers on Korean Island tourism. *Journal of Transport Geography* 19(6):1335–1340.

Coldren GM, Koppelman FS (2005) Modeling the competition among air-travel itinerary shares: GEV model development. *Transportation Research Part A: Policy and Practice* 39(4 SPEC. ISSS.):345–365.

Coldren GM, Koppelman FS, Kasturirangan K, Mukherjee A (2003) Modeling aggregate air-travel itinerary shares: logit model development at a major US airline. *Journal of Air Transport Management* 9(6):361–369.

Cordeau JF, Stojković G, Soumis F, Desrosiers J (2001) Benders decomposition for simultaneous aircraft routing and crew scheduling. *Transportation Science* 35(4):375–388.

Crouch GI (1994) Demand Elasticities for Short-Haul versus Long-Haul Tourism. *Journal of Travel Research* 33(2):2–7.

de Grange L, Fernández E, de Cea J (2010) A consolidated model of trip distribution. *Transportation Research Part E: Logistics and Transportation Review* 46(1):61–75.

de Grange L, Ibeas A, González F (2011) A Hierarchical Gravity Model with Spatial Correlation: Mathematical Formulation and Parameter Estimation. *Networks and Spatial Economics* 11(3):439–463.

Desaulniers G, Desrosiers J, Dumas Y, Solomon MM, Soumis F (1997) Daily aircraft routing and scheduling. *Management Science* 43(6):841–855.

Dobruszkes F (2006) An analysis of European low-cost airlines and their networks. *Journal of Transport Geography* 14(4):249–264.

Dong Z, Chuhang Y, Lau HY (2016) An integrated flight scheduling and fleet assignment method based on a discrete choice model. *Computers and Industrial Engineering* 98:195–210.

Dožić S, Kalić M (2015) Three-stage airline fleet planning model. *Journal of Air Transport Management* 46:30–39.

Driscoll JC, Kraay AC (1998) Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data. *Review of Economics and Statistics* 80(4):549–560.

Duran MA, Grossmann IE (1986) An outer-approximation algorithm for a class of mixed-integer nonlinear programs. *Mathematical Programming* 36(3):307–339.

Eronen VP, Mäkelä MM, Westerlund T (2015) Extended cutting plane method for a class of nonsmooth nonconvex MINLP problems. *Optimization* 64(3):641–661.

Exler O, Antelo LT, Egea JA, Alonso AA, Banga JR (2008) A Tabu search-based algorithm for mixed-integer nonlinear problems and its application to integrated process and control system design. *Computers and Chemical Engineering* 32(8):1877–1891.

Flyvbjerg B, Skamris Holm MK, Buhl SL (2005) How (In)accurate Are Demand Forecasts in Public Works Projects?: The Case of Transportation. *Journal of the American Planning Association* 71(2):131–146.

Forinash CV, Koppelman F (1993) Application and Interpretation of Nested Logit Models of Intercity Mode Choice. *Transportation Research Record* 1413.

Francis G, Dennis N, Ison S, Humphreys I (2007) The transferability of the low-cost model to long-haul airline operations. *Tourism Management* 28(2):391–398.

Freund-Feinstein U, Bekhor S (2017) An airline itinerary choice model that includes the option to delay the decision. *Transportation Research Part A: Policy and Practice* 96:64–78.

Furuichi M, Koppelman FS (1994) An analysis of air travelers' departure airport and destination choice behavior. *Transportation Research Part A* 28(3):187–195.

Gallego G, Ratliff R, Shebalov S (2015) A General Attraction Model and Sales-Based Linear Program for Network Revenue Management Under Customer Choice. *Operations Research* 63(1):212–232.

Gao C, Johnson E, Smith B (2009) Integrated airline fleet and crew robust planning. *Transportation Science* 43(1):2–16.

Garrow LA (2016) *Discrete Choice Modelling and Air Travel Demand: Theory and Applications.*

Garrow LA, Coldren GM, Koppelman FS, Coldren GM, Koppelman FS (2016) MNL, NL, and OGEV Models of Itinerary Choice 223–272.

Gayle PG (2013) On the efficiency of codeshare contracts between airlines: Is double marginalization eliminated? *American Economic Journal: Microeconomics* 5(4):244–273.

Geoffrion AM (1972) Generalized Benders decomposition. *Journal of Optimization Theory and Applications* 10(4):237–260.

Gil-Moltó MJ, Hole AR (2004) Tests for the consistency of three-level nested logit models with utility maximization. *Economics Letters* 85(1):133–137.

Greene WH, Hensher DA (2003) A latent class model for discrete choice analysis: Contrasts with mixed logit. *Transportation Research Part B: Methodological* 37(8):681–698.

Grosche T, Rothlauf F, Heinzl A (2007) Gravity models for airline passenger volume estimation. *Journal of Air Transport Management* 13(4):175–183.

Gronkvist M (2005) *The tail assignment problem* (Chalmers Univ. of Technology).

Guan W (2003) From the Help Desk: Bootstrapped Standard Errors. *The Stata Journal: Promoting communications on statistics and Stata* 3(1):71–80.

Guevara-Cue C (2010) *Endogeneity and sampling of alternatives in spatial choice models.* Ph.D. thesis, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA.

Gupta OK, Ravindran A (1985) Branch and bound experiments in convex nonlinear integer programming. *Management Science* 31(12):1533–1546.

Hakim MM, Merkert R (2016) The causal relationship between air transport and economic growth: Empirical evidence from South Asia. *Journal of Transport Geography* 56:120–127.

Hane CA, Barnhart C, Johnson EL, Marsten RE, Nemhauser GL, Sigismondi G (1995) The fleet assignment problem: Solving a large-scale integer program. *Mathematical Programming* 70(1-3):211–232.

Haouari M, Sherali HD, Mansour FZ, Aissaoui N (2011) Exact approaches for integrated aircraft fleeting and routing at TunisAir. *Computational Optimization and Applications* 49(2):213–239.

Hensher DA, Greene WH (2002) Specification and estimation of the nested logit model: alternative normalisations. *Transportation Research Part B: Methodological* 36(1):1–17.

Hensher DA, Greene WH (2003) The mixed logit model: The state of practice. *Transportation* 30(2):133–176.

Hensher DA, Rose JM, Greene WH (2005) *Applied choice analysis : a primer* (Cambridge University Press).

Hess S, Polak JW (2005) Mixed logit modelling of airport choice in multi-airport regions. *Journal of Air Transport Management* 11(2):59–68.

Hoechle D (2007) Robust Standard Errors for Panel Regressions with Cross-Sectional Dependence. *The Stata Journal: Promoting communications on statistics and Stata* 7(3):281–312.

Hsiao CY (2008) *Passenger demand for air transportation in a hub-and-spoke network.* Ph.D. thesis.

Hsiao CY, Hansen M (2011) A passenger demand model for air transportation in a hub-and-spoke network. *Transportation Research Part E: Logistics and Transportation Review* 47(6):1112–1125.

Hsu CI, Li HC, Liu SM, Chao CC (2011) Aircraft replacement scheduling: A dynamic programming approach. *Transportation Research Part E: Logistics and Transportation Review* 47(1):41–60.

Hsu CI, Wen YH (2000) Application of Grey theory and multiobjective programming towards airline network design. *European Journal of Operational Research* 127(1):44–68.

Hsu CI, Wen YH (2002) Reliability evaluation for airline network design in response to fluctuation in passenger demand. *Omega* 30(3):197–213.

Hsu CI, Wen YH (2003) Determining flight frequencies on an airline network with demand-supply interactions. *Transportation Research Part E: Logistics and Transportation Review* 39(6):417–441.

IATA (2019) World Air Transport Statistics 2019. Technical report.

Ishii J, Jun S, Van Dender K (2009) Air travel choices in multi-airport markets. *Journal of Urban Economics* 65(2):216–227.

Jacobs TL, Garrow LA, Lohatepanont M, Koppelman FS, Coldren GM, Purnomo H (2012) Airline planning and schedule development. *International Series in Operations Research and Management Science*, volume 169, 35–99 (Springer New York LLC).

Jaillet P, Song G, Yu G (1996) Airline network design and hub location problems. *Location Science* 4(3):195–212.

Jamili A (2017) A robust mathematical model and heuristic algorithms for integrated aircraft routing and scheduling, with consideration of fleet assignment problem. *Journal of Air Transport Management* 58:21–30.

Jankiewicz J, Huderek-Glapska S (2016) The air transport market in Central and Eastern Europe after a decade of liberalisation - Different paths of growth. *Journal of Transport Geography* 50:45–56.

Jaquillat A (2020) .

Jekel CF, Venter G (2019) PieceWise Linear Fitting: A Python Library for Fitting 1D Continuous Piecewise Linear Functions (March):1–15.

Jorge-Calderón J (1997) A demand model for scheduled airline services on international European routes. *Journal of Air Transport Management* 3(1):23–35.

Kasirzadeh A, Saddoune M, Soumis F (2017) Airline crew scheduling: models, algorithms, and data sets. *EURO Journal on Transportation and Logistics* 6(2):111–137.

Keha AB, De Farias IR, Nemhauser GL (2004) Models for representing piecewise linear cost functions. *Operations Research Letters* 32(1):44–48.

Kniker TS (1998) *Itinerary-based airline fleet assignment model.* Ph.d. thesis, MIT, Cambridge, MA.

Koppelman FS, Coldren GM, Parker RA (2008) Schedule delay impacts on air-travel itinerary demand. *Transportation Research Part B: Methodological* 42(3):263–273.

Koppelman FS, Wen CH (1998) Alternative nested logit models: structure, properties and estimation. *Transportation Research Part B: Methodological* 32(5):289–298.

Kummu M, Taka M, Guillaume JH (2018) Gridded global datasets for Gross Domestic Product and Human Development Index over 1990-2015. *Scientific Data* 5:1–15.

Kundu S, Ubhaya VA (2001) Fitting a least squares piecewise linear continuous curve in two dimensions. *Computers and Mathematics with Applications* 41(7-8):1033–1041.

Lee J, Wilson D (2001) Polyhedral methods for piecewise-linear functions I: The lambda method. *Discrete Applied Mathematics* 108(3):269–285.

Levin A (1971) Scheduling and Fleet Routing Models for Transportation Systems. *Transportation Science* 5(3):232–255.

Lhéritier A, Bocamazo M, Delahaye T, Acuna-Agost R (2018) Airline itinerary choice modeling using machine learning. *Journal of Choice Modelling* .

Li T, Wan Y (2019) Estimating the geographic distribution of originating air travel demand using a bi-level optimization model. *Transportation Research Part E: Logistics and Transportation Review* 131:267–291.

Liang Z, Chaovalitwongse WA (2013) A network-based model for the integrated weekly aircraft maintenance routing and fleet assignment problem. *Transportation Science* 47(4):493–507.

Lieshout R, Malighetti P, Redondi R, Burghouwt G (2016) The competitive landscape of air transport in Europe. *Journal of Transport Geography* 50:68–82.

Listes O, Dekker R (2005) A scenario aggregation-based approach for determining a robust airline fleet composition for dynamic capacity allocation. *Transportation Science* 39(3):367–382.

Lohatepanont M, Barnhart C (2004) Airline Schedule Planning: Integrated Models and Algorithms for Schedule Design and Fleet Assignment. *Transportation Science* 38(1):19–32.

Lurkin V, Garrow LA, Higgins MJ, Newman JP, Schyns M (2017) Accounting for price endogeneity in airline itinerary choice models: An application to Continental U.S. markets. *Transportation Research Part A: Policy and Practice* 100:228–246.

Lurkin V, Garrow LA, Higgins MJ, Newman JP, Schyns M (2018) Modeling competition among airline itineraries. *Transportation Research Part A: Policy and Practice* 113:157–172.

Malighetti P, Paleari S, Redondi R (2008) Connectivity of the European airport network: "Self-help hubbing" and business implications. *Journal of Air Transport Management* 14(2):53–65.

Margaretic P, Thomas-Agnan C, Doucet R (2017) Spatial dependence in (origin-destination) air passenger flows. *Papers in Regional Science* 96(2):357–380.

Marianov V, Serra D (2003) Location models for airline hubs behaving as M/D/c queues. *Computers and Operations Research* 30(7):983–1003.

Marla L, Vaze V, Barnhart C (2018) Robust optimization: Lessons learned from aircraft routing. *Computers and Operations Research* 98:165–184.

McFadden D (1974) The measurement of urban travel demand. *Journal of Public Economics* 3(4):303 – 328.

McGill JI, Van Ryzin GJ (1999) Revenue management: research overview and prospects. *Transportation Science* 33(2):233–256.

Mercier A, Cordeau JF, Soumis F (2005) A computational study of Benders decomposition for the integrated aircraft routing and crew scheduling problem. *Computers and Operations Research* 32(6):1451–1476.

Mercier A, Soumis F (2007) An integrated aircraft routing, crew scheduling and flight retiming model. *Computers and Operations Research* 34(8):2251–2265.

Mokhtarian PL (2016) Presenting the independence of irrelevant alternatives property in a first course on logit modeling. *Journal of Choice Modelling* 21:25 – 29, standalone technical contributions in choice modelling.

Mumbower S, Garrow LA, Higgins MJ (2014) Estimating flight-level price elasticities using online airline data: A first step toward integrating pricing, demand, and revenue optimization. *Transportation Research Part A: Policy and Practice* 66(1):196–212.

O'Kelly ME (1987) A quadratic integer program for the location of interacting hub facilities. *European Journal of Operational Research* 32(3):393–404.

Ortuzar JdD, Willumsen LG (2011) *Modelling transport* (Wiley-Blackwell).

Padberg M (2000) Approximating separable nonlinear functions via mixed zero-one programs. *Operations Research Letters* 27(1):1–5.

Paleari S, Redondi R, Malighetti P (2010) A comparative study of airport connectivity in China, Europe and US: Which network provides the best service to passengers? *Transportation Research Part E: Logistics and Transportation Review* 46(2):198–210.

Papadakos N (2009) Integrated airline scheduling. *Computers and Operations Research* 36(1):176–195.

Park Y, Ha HK (2006) Analysis of the impact of high-speed railroad service on air transport demand. *Transportation Research Part E: Logistics and Transportation Review* 42(2):95–104.

Pels E, Nijkamp P, Rietveld P (2003) Access to and competition between airports: a case study for the San Francisco Bay area. *Transportation Research Part A: Policy and Practice* 37(1):71–83.

Pita JP, Adler N, Antunes AP (2014) Socially-oriented flight scheduling and fleet assignment model with an application to Norway. *Transportation Research Part B: Methodological* 61:17–32.

Pita JP, Barnhart C, Antunes AP (2013) Integrated Flight Scheduling and Fleet Assignment Under Airport Congestion. *Transportation Science* 47(4):477–492.

Quesada I, Grossmann IE (1992) An LP/NLP based branch and bound algorithm for convex MINLP optimization problems. *Computers and Chemical Engineering* 16(10-11):937–947.

Repko MG, Santos BF (2017) Scenario tree airline fleet planning for demand uncertainty. *Journal of Air Transport Management* 65:198–208.

Rexing B, Barnhart C, Kniker T, Jarrah A, Krishnamurthy N (2000) Airline Fleet Assignment with Time Windows. *Transportation Science* 34(1):1–20.

Rubin DB (1976) Inference and missing data. *Biometrika* 63(3):581–592.

Sa CA, Santos BF, Clarke JPB (2019) Portfolio-based airline fleet planning under stochastic demand. *Omega (United Kingdom)* 102101.

Sandhu R, Klabjan D (2007) Integrated airline fleeting and crew-pairing decisions. *Operations Research* 55(3):439–456.

Saunders JA, Morrow-Howell N, Spitznagel E, Dore P, Proctor EK, Pescarino R (2006) Imputing Missing Data: A Comparison of Methods for Social Work Researchers. *Social Work Research* 30(1):19–31.

Schlüter M, Egea JA, Banga JR (2009) Extended ant colony optimization for non-convex mixed integer nonlinear programming. *Computers and Operations Research* 36(7):2217–2229.

Shao J, Sitter RR (1996) Bootstrap for Imputed Survey Data. *Journal of the American Statistical Association* 91(435):1278–1288.

Sherali HD, Bae KH, Haouari M (2010) Integrated Airline Schedule Design and Fleet Assignment: Polyhedral Analysis and Benders' Decomposition Approach. *INFORMS Journal on Computing* 22(4):500–513.

Sherali HD, Bae KH, Haouari M (2013) A benders decomposition approach for an integrated airline schedule design and fleet assignment problem with flight retiming, schedule balance, and demand recapture. *Annals of Operations Research* 210(1):213–244.

Sherali HD, Bish EK, Zhu X (2005) Polyhedral Analysis and Algorithms for a Demand-Driven Refleeting Model for Aircraft Assignment. *Transportation Science* 39(3):349–366.

Sherali HD, Bish EK, Zhu X (2006) Airline fleet assignment concepts, models, and algorithms. *European Journal of Operational Research* 172(1):1–30.

Smyth M, Pearce B (2008) IATA Economics Briefing N9: Air Travel Demand. Technical report.

Soylu B, Katip H (2019) A multiobjective hub-airport location problem for an airline network design. *European Journal of Operational Research* 277(2):412–425.

Still C, Westerlund T (2008) Extended Cutting Plane Algorithm. *Encyclopedia of Optimization*, 959–966 (Springer US).

Storn R, Price K (1997) Differential Evolution – A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces. *Journal of Global Optimization* 11:341–359.

Strauss AK, Klein R, Steinhardt C (2018) A review of choice-based revenue management: Theory and methods. *European Journal of Operational Research* 271(2):375–387.

Subramanian R, Scheff RP, Quillinan JD, Wiper DS, Marsten RE (1994) Coldstart: Fleet Assignment at Delta Air Lines. *Interfaces* 24(1):104–120.

Swan WM, Adler N (2006) Aircraft trip cost parameters: A function of stage length and seat capacity. *Transportation Research Part E: Logistics and Transportation Review* 42(2):105–115.

Talluri K, van Ryzin G (2004) Revenue Management Under a General Discrete Choice Model of Consumer Behavior. *Management Science* 50(1):15–33.

Tatem AJ (2017) WorldPop, open data for spatial demography. *Scientific Data* 4(1):1–4.

Teodorović D, Kalić M, Pavković G (1994) The potential for using fuzzy set theory in airline network design. *Transportation Research Part B* 28(2):103–121.

Train KE (2003) *Discrete choice methods with simulation* (Cambridge: Cambridge University Press).

Valdes V (2015) Determinants of air travel demand in Middle Income Countries. *Journal of Air Transport Management* 42:75–84.

Vaze V, Barnhart C (2012) Modeling airline frequency competition for airport congestion mitigation. *Transportation Science* 46(4):512–535.

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

Vielma JP, Nemhauser GL (2011) Modeling disjunctive constraints with a logarithmic number of binary variables and constraints. *Mathematical Programming* 128(1-2):49–72.

Wei K, Vaze V, Jacquillat A (2020) Airline Timetable Development and Fleet Assignment Incorporating Passenger Choice. *Transportation Science* 54(1):139–163.

Wei W, Hansen M (2005) Impact of aircraft size and seat availability on airlines' demand and market share in duopoly markets. *Transportation Research Part E: Logistics and Transportation Review* 41(4):315–327.

Wei W, Hansen M (2006) An aggregate demand model for air passenger traffic in the hub-and-spoke network. *Transportation Research Part A: Policy and Practice* 40(10):841–851.

Wen CH, Lai SC (2010) Latent class models of international air carrier choice. *Transportation Research Part E: Logistics and Transportation Review* 46(2):211–221.

Wen YH, Hsu CI (2006) Interactive multiobjective programming in airline network design for international airline code-share alliance. *European Journal of Operational Research* 174(1):404–426.

Westerlund T, Pettersson F (1995) An extended cutting plane method for solving convex MINLP problems. *Computers and Chemical Engineering* 19(SUPPL. 1):131–136.

Westerlund T, Skrifvars H, Harjunkoski I, Pörn R (1998) An extended cutting plane method for a class of non-convex MINLP problems. *Computers and Chemical Engineering* 22(3):357–365.

Wojahn OW (2001) Airline network structure and the gravity model. *Transportation Research Part E: Logistics and Transportation Review* .

Xiao Y, Liu Y, Liu JJ, Xiao J, Hu Y (2016) Oscillations extracting for the management of passenger flows in the airport of Hong Kong. *Transportmetrica A: Transport Science* 12(1):65–79.

Yan C, Barnhart C, Vaze V (2020) Choice-Based Airline Schedule Design and Fleet Assignment: A Decomposition Approach. *SSRN Electronic Journal* .

Yang TH (2010) A two-stage stochastic model for airline network design with uncertain demand. *Transportmetrica* 6(3):187–213.