Airline long-haul network development practices: an extensive worldwide empirical investigation

Andrea Signori[a], Nicolò Avogadro[a,*], Sebastian Birolini[a,b]

[a] *Department of Management, Information and Production Engineering, University of Bergamo, Dalmine (BG) Italy.*
[b] *flAi S.r.l., Curno (BG) Italy.*

Abstract

Airline network development and new-route selection are complex strategic processes that play a critical role in shaping airline performance and competitive positioning. Despite their importance, systematic and scalable analytical tools to support route-entry decisions remain limited, particularly in long-haul markets. This paper proposes a data-driven classification model to identify promising long-haul route-entry opportunities based on publicly available data. The model integrates features capturing market potential, competitive intensity, fleet capacity, and feeding potential, and is applied to a global dataset of nonstop long-haul routes over the period 2014–2019. Given the sporadic nature of route-entry events and the resulting class imbalance, alternative resampling strategies are evaluated, with oversampling yielding more stable results. Carrier-specific logit models are compared with a pooled specification, revealing substantial heterogeneity in the factors affecting network development strategies across airlines. Connecting demand consistently emerges as a key factor, while the relative importance of socioeconomic conditions, fleet capacity, and feeding potential varies across carriers. The practical relevance of the approach is illustrated through case studies of major airlines, where several routes identified as promising were subsequently launched, demonstrating the model's usefulness as a pre-screening tool for airline network planning.

Keywords: Airline network development; Network planning; Long-haul routes; Classification model; Data-driven decision-making.

1. Introduction

The exponential growth of airline networks in recent decades has significantly contributed to the economies of cities and countries. Central to this dynamic is the strategic expansion of airline route networks. By strategically adding new routes and enhancing connectivity, airlines continually improve their service quality, ensuring that most passengers can access their desired destinations within the airline's network or through alliance partners (Wong et al. 2023). However, launching new routes represents a strategic investment that extends beyond mere geographic expansion. By constantly redesigning their networks, airlines aim to preserve their competitiveness, strengthen their market position, and augment the overall value of their route portfolios.

Airlines continuously evaluate how the introduction of new routes or modifications to existing ones—whether by themselves or competitors—impact their networks and commercial performance.

*Corresponding author

Email addresses: `andrea.signori1@unibg.it` (Andrea Signori), `nicolo.avogadro@unibg.it` (Nicolò Avogadro), `sebastian.birolini@unibg.it` (Sebastian Birolini)

As global connectivity expands, the addition of a new route can influence not only the specific market where the connection is introduced but also, through cascading effects, other markets that the new leg may serve. Additionally, a new route can significantly alter existing passenger flows by shifting transit options to different hubs or attracting the demand previously served through connecting itineraries. These complex network effects—which are particularly pronounced in long-haul markets—combined with fleet capacity constraints and the pursuit of strategic objectives, make airline network development far from trivial. In this regard, previous research offers no clear consensus on the factors influencing the choice of opening a new route and their relative importance within the airline decision-making process.

Despite the strategic relevance of new route selection decisions, data-driven tools to systematically analyze them remain scarce. In practice, most airlines still rely on case-specific market analyses or simplistic route-profitability models (Bannò and Redondi 2014, Halpern and Graham 2015, Carmona-Benítez et al. 2017). Promising decision-support tools include route network optimization models, which can assist airlines in determining which routes to launch or discontinue and which markets to prioritize. These provide superior benefits compared to traditional methods because of their ability to simultaneously consider multiple factors and effectively assess how each route fits within the overall airline network. However, airlines face substantial challenges in extensively employing network-planning algorithms because of their high computational demands. The main concern is the combinatorial explosion of potential itineraries; that is, as the number of airports and potential routes grows, the number of possible connections increases exponentially. This results in a vast search space that is computationally challenging to navigate exhaustively, limiting the applicability of such approaches to most real-world scenarios. At the same time, scholarly attention to analytics-based solutions for identifying new route-development opportunities is relatively sparse, with most studies addressing combinatorial complexity by considering relatively small-scale networks (Hausladen and Schosser 2020, Birolini et al. 2021).

In this paper, we propose a classification model to identify promising new long-haul routes based on publicly available data, a segment characterized by strong network effects and high strategic relevance for airlines. The model provides a comprehensive examination of the multifaceted factors and dynamics that influence airline route-entry decisions. Furthermore, it can be used to evaluate the network development potential of individual carriers and allows scalability to be integrated with existing network optimization algorithms to refine route selection strategies.

In detail, we formulate a classification logit model designed to systematically compare the characteristics of newly launched routes with potential ones. We leverage a comprehensive dataset of new routes opened between 2014 and 2019 on a global scale and implement a tailored procedure to build instances of potential routes that could have been opened but were not. We engineer features to capture key drivers—including market potential, competitive intensity, and network dynamics—to uncover the fundamental factors influencing airlines' network development decisions. Given the highly unbalanced nature of the dataset due to the relative scarcity of new route openings, we test different resampling procedures, ultimately selecting the most robust approach. By proposing carrier-specific formulations, we conduct a cross-carrier comparison to investigate the pivotal factors influencing route decisions. We find substantial heterogeneity among carriers, which explains why different carriers employ different route-planning strategies. The model's performance is validated through tailored out-of-sample testing, showing variations in key performance indicators as different probability threshold values are considered. Finally, we demonstrate the model's real-world applicability in supporting network planning through a set of case studies. In these applications, the model predicts new route openings based on 2019 data and compares them with the routes actually launched by airlines in subsequent years. The results underscore the model's predictive accuracy and its practical utility in selecting promising routes for further investigation within network-planning algorithms or for strategic and operational evaluation.

The remainder of the paper is organized as follows. Section 2 discusses the relevant literature. Section 3 details the sample identification, features engineering, and methodology. Section 4 presents the empirical results and introduces evidence from the tailored out-of-sample validation experiment. Finally, Section 5 concludes and provides avenues for future research.

2. Literature review

2.1. Route-entry decisions

There is substantial research examining airlines' route-entry decisions from an empirical perspective. The aim of these contributions is to demonstrate the drivers underpinning airlines' route-entry decisions by *ex post* analyzing the market characteristics and competitive environments in which such events occur. Following the seminal study by Morrison and Winston (1990), the key drivers influencing route-entry decisions have been categorized into socioeconomic factors, market competition, and network considerations (Abdelghany and Guzhva 2010, Halpern and Graham 2015, Hanson et al. 2022). Socioeconomic factors include variables such as population size, income levels, and regional economic growth, which ultimately determine the potential market size for a given route. Market competition reflects the competitive pressure within the market and depends on the number of competitors, their fare structure, and aspects that collectively influence the attractiveness of entering the market. Finally, network considerations encompass a broad spectrum of factors related to the role of hub connectivity, route density, and market coverage. These are often airline- or airline group-specific factors and are shaped by airline strategy and market coverage—for example, a geographical or market-segment focus—subject to fleet availability constraints (Oliveira 2008, Zhang et al. 2017).

Despite numerous studies analyzing route-entry decisions, there is no clear consensus on the factors affecting route-entry choices and their relative importance within the airline decision-making process. More importantly, prior studies have typically adopted a case-based approach focusing on single airlines or specific geographical contexts, such as single countries (Fu et al. 2015, Calzada and Fageda 2019, Gaggero and Piazza 2021), likely because of the limited data available. These studies—often regional case studies (see Dresner et al. 2015, Wang et al. 2017, Zhang et al. 2017, Wang et al. 2022) or carrier-specific case studies (see Boguslaski et al. 2004, Aydemir 2012, Fu et al. 2015, Zou and Yu 2020)— have typically aimed to explain specific situations rather than uncover broader, generalizable results. As a result, they have failed to identify overarching patterns or trends that apply across different markets or airlines. The limited ability to collect relatively large datasets has also constrained most studies to short-haul markets, in which seasonal connections are more common and market conditions are more dynamic (Boguslaski et al. 2004, Zou and Yu 2020).[1] The *ex post* evaluation approach has also limited

[1]The stronger presence of low-cost companies in this travel segment facilitates the collection of larger datasets due to

the analysis of factors determining airlines' entry into established markets rather than investigating and systematically comparing the characteristics of first-time-served markets—that is, markets not previously served by nonstop services—against the entire set of potential routes. Only recently have some studies targeted the determinants of first-time-served markets, focusing on their managerial implications (Abdelghany and Guzhva 2022, Wong et al. 2023). However, these contributions still rely on limited samples and do not provide a scalable framework for systematically screening potential long-haul route-entry opportunities across carriers.

In summary, although existing studies have provided valuable insights into carrier-specific dynamics or specific geographic and market contexts, their findings remain difficult to generalize beyond the settings analyzed. The prevailing reliance on case-based and short-haul-focused analyses has limited the ability to capture the broader strategic and network-driven considerations that underpin airlines' route-entry decisions across heterogeneous market environments. This limitation is particularly evident for long-haul connections, which *de facto* still constitute the key pillar of hub-and-spoke systems and the core of global connectivity (Lee et al. 2014, Cheung et al. 2022).

2.2. Strategic network planning

Route planning is a core element of the airline planning process. This process involves a complex interplay of strategic, tactical, and operational decisions aimed at maximizing airline profitability while achieving business goals and maintaining the required flexibility. While tactical and operational decisions play a pivotal role in optimizing resource allocation and promoting operational efficiency in the short to medium term (Zografos et al. 2012, Kim and Hansen 2015), strategic planning is fundamental for long-term growth and competitive positioning, ultimately shaping an airline's overall success in the market (Mohri et al. 2022, Wu et al. 2022, Geursen et al. 2023). These strategic choices are directly tied to network development and influence how airlines build, expand, and refine the routes they operate.

Research has devoted significant attention to developing analytical models that support tactical and operational decisions. Relevant examples include contributions focusing on flight scheduling and fleet assignment, crew scheduling, aircraft routing, and recovery decisions (Santana et al. 2023, Xu et al. 2024, Wen et al. 2024). In contrast, strategic network planning has received less attention. In this domain, most academic research has focused on hub location and fleet planning problems (Mohammadi et al. 2019, Soylu and Katip 2019, Alumur et al. 2021). Research in the former area involves determining the location of hub airports to connect passenger flows from origin to destination nodes by modeling network architectures—hub & spoke versus point-to-point—that maximize profits or minimize costs (Alumur et al. 2021, Sharma et al. 2024). Despite their important theoretical insights, these models are of more limited managerial relevance, since hub locations generally remain fixed for most carriers from their inception. In contrast, fleet planning has been explored through optimization models that support decisions on when to dismiss or sell aircraft and when to acquire or lease them, considering factors such as demand uncertainty, purchasing and leasing alternatives, and operating costs (Teoh and Khoo 2016, Baykasoğlu et al. 2022).

Besides hub locations and fleet planning, as part of their long-term strategy, airlines need to frequently evaluate which routes to open, close, or connect. As currently practiced, most airlines

their practice of rotating leisure destinations.

identify which market to enter based on market analysis, basic econometric models estimating potential demand, or simplistic route profitability models (Bannò and Redondi 2014, Halpern and Graham 2015, Carmona-Benítez et al. 2017). Despite the critical importance of these decisions in shaping airline networks, route planning has received limited attention in the academic literature, with only a few models tackling airline network expansion design (Kölker and Lütjens 2015, Birolini et al. 2021). The main reason for this is the high computational cost of network-planning algorithms, which makes it difficult to apply these models to real-world scenarios (Carreira et al. 2017, Schosser and Schosser 2020). This follows from the combinatorial complexity inherent in network planning, in which the vast number of potential route combinations makes it challenging to develop efficient algorithms capable of identifying optimal network structures. The few studies that addressed this issue attempted to tackle the combinatorial complexity, but this presented a major challenge even for relatively small-scale networks (Teodorović et al. 1994, Jaillet et al. 1996). In general, current approaches to route planning and network expansion design often lack a preliminary model that can narrow down the pool of potential routes before applying optimization algorithms or to be used as input for kernel search heuristics in large-scale network optimization problems. Such a model could provide valuable insights into identifying the most promising candidates for network expansion, reducing the search space, and making the optimization problem computationally more tractable. By integrating this type of preprocessing tool, network-planning algorithms would operate more efficiently, significantly cutting down on computational time, thus paving the way for their widespread adoption within airline decision-making processes.

In short, while airlines have extensively leveraged advanced optimization tools for tactical and operational decisions, analytics-based solutions for strategic network development—particularly route planning—remain relatively scarce, despite their potential to deliver significant benefits and valuable insights.

This paper addresses these gaps by proposing a classification model to identify promising long-haul routes based on a set of publicly available variables. The model, applied to new routes opened between 2014 and 2019 worldwide, provides a comprehensive empirical assessment of the determinants of route openings and offers insights into airlines' network expansion decisions. Moreover, we elaborate on the potential use of this approach within airline decision-support tools, either as a complement to network optimization models or as a preliminary screening instrument for identifying potential destinations for further investigation.

3. Research design

This study applies a classification model to empirically investigate how airlines decide to enter new long-haul markets. This model leverages a comprehensive dataset of nonstop long-haul routes opened between 2014 and 2019 worldwide. The characteristics of these routes are compared against those of routes that could potentially have been opened. We engineer a set of explanatory variables to capture essential drivers influencing airlines' network development decisions, including market potential, competition, and network dynamics. To investigate the effects of such variables, we apply a classification model with carrier-specific formulations to analyze heterogeneous airlines' behaviors in selecting new markets for entry.

Table 1: Number of observations by year and type for the top 20 carriers by long-haul routes operated in 2019.

Year	Canceled	Established	New
2014	108	2750	306
2015	135	2921	239
2016	119	3052	321
2017	149	3224	281
2018	104	3401	337
2019	173	3565	296
Total	788	18,913	1780

In this section, we introduce the sample-building procedure, the variables considered, and the methodology.

3.1. Sample identification

The data required for the proposed approach are twofold: (i) information on airline entry—that is, new long-haul connections launched by selected airlines during the study period—and (ii) potential candidate routes to be operated.

To gather data on route entry, the OAG Schedule Analyser database was used. The period between 2014 and 2019 was examined, and route-level data for each carrier worldwide were collected. Each route was uniquely identified by the connected airports—namely, airport pairs—and the operating airline. This led to a dataset comprising more than 600,000 route-carrier-year observations. To focus on long-haul markets, we applied a fixed cutoff for a route great-circle distance of 3000 km. We further restricted our sample to observations with at least 25 flights per year, assuming it to be a reasonable threshold for identifying active connections corresponding to at least seasonal weekly services. From this initial dataset, we restrict the empirical analysis to the top 20 full-service carriers by number of long-haul routes operated in 2019, reflecting their dominant role in long-haul markets and their extensive use of hub-and-spoke network structures.

At this point, we classified each observation (route-carrier-year) as *established* if the route was operated both in year t and in the previous year by the single airline; *new* if the route was launched by the airline in year t; and *canceled* if the route ceased operation by the airline in year t but was operated in the previous year. Table 1 summarizes the number of observations based on the year and type for the top 20 carriers by number of long-haul routes operated in 2019. As expected, we observed that the majority of long-haul routes were marked as established. This subgroup accounted for about 2750 route-carrier pairs in 2014 and increased to 3565 in 2019. The number of new long-haul routes ranged from 239 to 337 per year during the period under investigation. Finally, the number of canceled routes increased from 108 in 2014 to 173 in 2019. In our analysis, we focus on the route-carrier observations marked as new.[2] These are routes in which a specific airline entered in a given year.

Newly opened routes needed to be compared with those that could have potentially been launched by airlines. We built a tailored procedure to generate such observations. Given the focus on full-

[2]Due to peculiar market-entry determinants, we exclude from our analysis flights operated exercising 5^{th} Freedom rights—that is, flights between two countries that are not the carrier's home country.

service carriers—typically operating long-haul flights from their hub facilities—we first identified the airports that could be considered hubs for each airline. For this classification, we referred to the historical number of connecting passengers sourced from OAG Traffic Analyser and designated as hubs the airports where each airline managed the majority of its connecting passengers. Most of the airlines considered had only one airport meeting the criteria, which was identified as their hub (e.g., DXB for Emirates and LHR for British Airways), although a few airlines had a multi-hub structure (e.g., Lufthansa with FRA and MUC). This ultimately demonstrates the threshold's capability to identify each carrier's network structure (single- vs. multi-hub). Second, we identified potential candidate destination airports by considering all airports worldwide with scheduled long-haul flights, using this as a proxy for the technical infrastructure capability to accommodate such traffic, such as runway length and terminal/stand infrastructure. To reasonably restrict the potential destinations, we applied filtering criteria excluding airports with a number of historical long-haul movements below the 20^{th} percentile for each geographical macroregion. This sample restriction allowed us to retain only airports with consistent long-haul traffic and that could be considered suitable potential destinations. At this point, we identified the set of potential routes for each airline as those connecting airline hubs with the candidate destination airports, and *vice versa*. To maintain the focus on long-haul connections, we applied a distance cutoff by selecting only potential routes with distances greater than 3000 km. Moreover, to ensure the technical feasibility of such flights, we restricted the analysis to potential routes below 12,545 km—that is, the 99^{th} percentile of the distance of existing long-haul routes. Finally, we excluded the observations (route-carrier-year) already observed in the historical dataset (as established, new, or canceled) from the set of potential routes.

The final dataset comprised the new routes—namely, routes that were opened by an airline in a specific year—and potential route candidates each airline could have operated but did not. To preserve results tractability while maintaining the global scope of the analysis, for the empirical analysis, we focus on observations of the top 20 carriers based on the number of long-haul routes operated in 2019. Table 2 reports the number of observations (route-carrier-year) by carrier and type. The final sample consisted of 87,999 observations: 1780 new routes and 86,219 potential routes.

3.2. Feature engineering

In this section, we introduce and discuss the formulation of explanatory variables designed to capture the essential drivers influencing airline network development. We begin engineering and testing a broad set of features across various combinations within the empirical model. Through a comprehensive evaluation, we assess the model's performance using different subsets of variables to determine their impact. After carefully considering the results, we retain only the most relevant variables—namely, those contributing significantly to the model's predictive power without introducing multicollinearity issues. While the final set of explanatory variables used in the empirical models is specified in Section 4.1, this section presents the entire set of explanatory variables considered.

To systematically capture the key drivers influencing airlines' network development decisions, we consider five groups of factors grounded in prior research: route distance, capacity constraints, market potential, market competition, and feeding potential. Table 3 summarizes these categories along with the definitions of the variables analyzed within each group. Figure 1 illustrates the correlations between the different variables.

Table 2: Sample by carrier and type of observation: New or potential.

Carrier	IATA Code	Route-carrier-year observations		
		New	**Potential**	**Total**
Delta Airlines	DL	190	5525	5715
Air Canada	AC	183	9544	9727
United Airlines	UA	128	8469	8597
Alaska Airlines	AS	122	458	580
China Southern Airlines	CZ	120	2818	2938
China Eastern Airlines	MU	116	6146	6262
American Airlines	AA	115	5350	5465
Qatar Airways	QR	99	2466	2565
Air China	CA	98	3098	3196
Turkish Airlines	TK	83	2501	2584
Saudi Arabian Airlines	SV	81	6637	6718
Lufthansa	LH	68	5777	5845
Aeroflot	SU	60	3094	3154
LATAM Airlines	LA	59	7948	8007
Ethiopian Airlines	ET	58	3215	3273
British Airways	BA	52	2666	2718
Emirates	EK	48	2285	2333
Air France	AF	45	2536	2581
KLM Royal Dutch Airlines	KL	37	2812	2849
Korean Air	KE	18	2874	2892
Total		1780	86,219	87,999

The first dimension considered is distance. Route distance represents the most straightforward measure of travel impedance considered within basic gravity model frameworks, which assume that economic connectivity between regions and travel demand decreases as distance increases. This weaker coupling is also expected to affect the attractiveness of serving a market and thus opening a new route. Accordingly, in our analysis, we consider the great-circle distance between the origin and destination airports, calculated based on their geographical coordinates.

The second key dimension in assessing the feasibility of opening a new route is fleet availability. In their planning process, airlines determine routes to operate based on the availability of aircraft with different capacities and operating ranges. To capture this aspect, we introduce a carrier-level capacity proxy based on year-over-year changes in available seat kilometers (ASK). Specifically, for each airline and year, we compute the variation in total ASK within three distance categories (3000–4000 km, 4000–6000 km, and over 6000 km). To account for the heterogeneous use of aircraft across distance segments, ASK is calculated considering only aircraft models typically deployed by the airline on routes of comparable length, defined as those falling between the 10^{th} and 90^{th} percentiles of seating capacity. The resulting ΔASK reflects changes in fleet capacity within each distance segment and is assigned to all new and potential routes operated by the airline in the corresponding distance category. Positive values indicate fleet expansion and potential new route openings, while negative values may indicate capacity reductions or tighter fleet constraints.

The third category of variables affecting route openings is socioeconomic factors, which ultimately shape the potential market size. In our analysis, we consider the population and GDP per capita

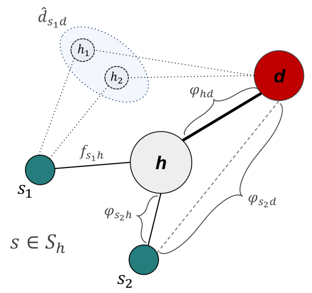
Category	Variable	Description and formulation
Distance	Route distance	Route great-circle distance
Capacity	Δ ASK	Year-over-year change in available seat kilometers (ASK), used as a proxy for fleet capacity. The measure is computed at the airline level for each flight distance category (3000–4000 km, 4000–6000 km, and over 6000 km) and based on aircraft models typically deployed on routes of comparable length (defined by the 10^{th}-90^{th} percentile of seating capacity).
Market Potential	Population	Total population within the destination airport catchment area, assumed as a static 100 km radius area
	GDP per capita	Average GDP per capita for inhabitants living within the destination airport catchment area, assumed as a static 100 km radius area
	CAGR population	Compound annual growth rate (CAGR) over a 5-year period of the population within the destination airport catchment area
	CAGR GDP per capita	Compound annual growth rate over a 5-year period of GDP per capita within the destination airport catchment area
	Nonstop demand	Number of passengers traveling between the departure city and the destination city using nonstop services
	Connecting demand	Number of passengers traveling between the departure city and the destination city using one-stop itineraries
Market Competition	No.airlines	Number of airlines operating the route
	No. partner	Number of airlines operating the route that belong to the same alliance as the carrier under consideration
	No. competitors	Number of airlines operating the route that do not belong to the same alliance as the carrier under consideration
	$FTSM$	First-time-served market: Dummy variable equal to 1 if the market is not served by any nonstop service and 0 otherwise
	HHI	Herfindahl-Hirschman Index based on the nonstop seating capacity. Let $\mathcal{A}$ be the set of airlines offering nonstop services on a route and ϑ_a the number of allocated seats by airline a; we define the HHI as $\sum_{a\in\mathcal{A}} MS_a^2$ where $MS_a = \frac{\vartheta_a}{\sum_{a\in\mathcal{A}}\vartheta_a}$ is the market share of airline a in terms of offered seats.
Feeding Potential	Equivalent number of spokes	For a given long-haul route from hub h to destination d, let S_h be the set of spokes s ($s \in S_h$) connected to the hub with at least one weekly flight. We define the equivalent number of spokes for the route from h to d as $SE_{hd} = \sum_{s\in\mathcal{S}_h} \frac{1}{Rf_{shd}}$, where Rf_{shd} is the routing factor, defined as $Rf_{shd} = \frac{\varphi_{sh}+\varphi_{hd}}{\varphi_{sd}}$ where φ represents the great-circle distance. 
	Corrected frequency	We define the corrected frequency by considering the equivalent number of spokes and adjusting the weight associated with each spoke s using the feeding frequency f_{sh} (the frequency of connections from the spoke to the hub). Accordingly, the corrected frequency is defined as $CF_{hd} = \sum_{s\in\mathcal{S}_h} \frac{f_{sh}}{Rf_{shd}}$, where Rf_{shd} is the routing factor.
	Corrected potential demand	Let $\hat{d}_{sd}$ be the number of passengers currently traveling from spoke s to final destination d using one-stop itineraries; we define the corrected potential demand as $CD_{hd} = \sum_{s\in\mathcal{S}_h} \frac{\hat{d}_{sd}}{Rf_{shd}}$, where Rf_{shd} is the routing factor.

Table 3: Variable definitions and formulations.

of the destination area.[3] To compute these values, we assume a static catchment area with a 100 km radius around each destination airport and derive the population and GDP values based on a high-resolution global spatial dataset (for details, see Tatem 2017, Kummu et al. 2018). These metrics offer insights into economic activity and consumer demand within the airport's vicinity, serving as key indicators of market size and purchasing power—both of which are critical factors in determining route attractiveness for airlines. In addition to the current socioeconomic figures, we also account for long term trends by incorporating the compound annual growth rate of population and gross domestic product (GDP) per capita over a five-year period, thereby capturing prospective regional demographic and economic growth.

Besides socioeconomic factors contributing to determining market size, market potential can also be assessed by examining current air demand. Accordingly, we consider both nonstop and connecting demand retrieved from OAG Traffic Analyser on routes under investigation. These two measures provide distinct yet complementary insights into route attractiveness. Nonstop demand refers to the volume of passengers using nonstop services on a particular market and results in a twofold interpretation of route attractiveness. On the one hand, high volumes of nonstop traffic indicate a strong and well-established market, suggesting that the route serves a thriving market. On the other hand, it may also imply that the market is already well-served by existing nonstop services, potentially leaving less room for new entrants. In contrast, connecting demand represents the number of passengers traveling on the analyzed market using one-stop itineraries. The higher this demand, the greater the number of passengers who could benefit from a more convenient nonstop service. A strong presence of connecting demand is indeed a signal of market potential—if a substantial number of passengers are already traveling between two cities via one-stop itineraries, introducing a more convenient nonstop flight could capture this latent demand, positioning the new route as an attractive alternative to existing connecting itineraries.

The attractiveness of entering a new route also depends on the competitive landscape. Accordingly, we consider the number of competing airlines already operating the route as a variable that may influence its attractiveness. At the same time, especially on long-haul routes, airlines often offer code-share opportunities. To account for this, we incorporate alliance dynamics by considering the number of partners already serving the route. Airlines within the same alliance are classified as partners, while those outside the alliance are considered competitors. This distinction provides insights into the competitive dynamics within the market as well as the carrier's positioning relative to both its alliance partners and competitors when evaluating the possibility of serving a new route. Besides the number of airlines, we also consider the overall competition in the market using the Herfindahl-Hirschman Index (HHI), computed based on allocated seating capacity retrieved from OAG Schedule Analyser. Ultimately, we model the entrance to a first-time-served market by using a dummy variable equal to 1 if the market is not served by nonstop services by any airline.

Finally, route-entry decisions also depend on airline network considerations, including how the potential route fits within the overall airline network. To cover this, we develop three metrics aimed at measuring the strength of the airline's feeder network for serving a specific long-haul route. Since these measures strictly relate to airline hubbing activities and the possibility of connecting different

[3]Given that the carrier home country and hub airport can reasonably be assumed to be fixed, we consider only socioeconomic variables regarding the route destination.

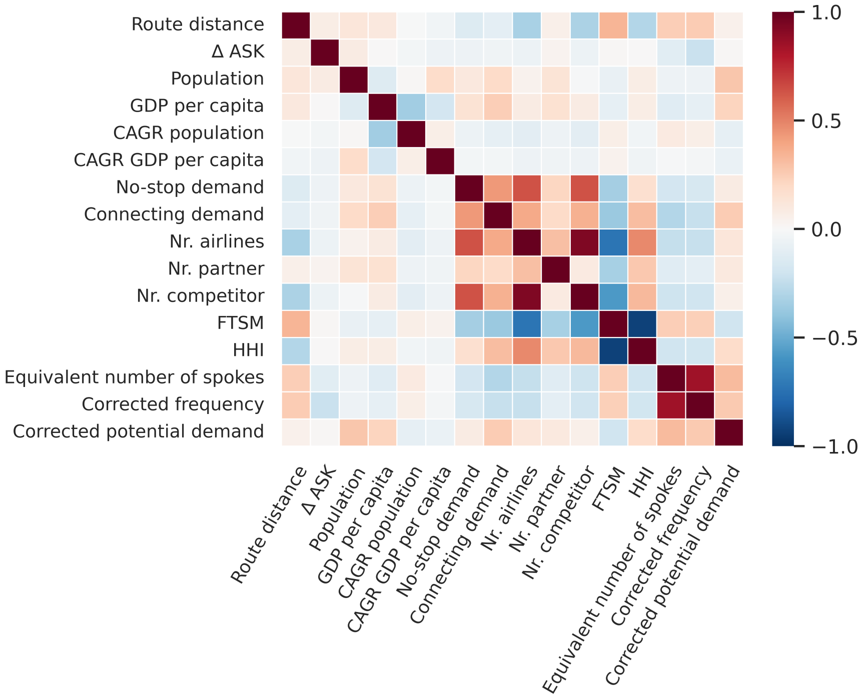


Figure 1: Heatmap of the correlation between variables (train dataset balanced through oversampling).

flights, we define these measures only for routes to and from the airline hub airport. The first measure is the equivalent number of spokes, which is based on the number of airports with which the hub airport is connected other than the long-haul destination considered . These airports, which we define as spokes, constitute the potential feeding base for the long-haul flight considered. For a given long-haul route from hub h to destination d, let us define S_h as the set of spokes s ($s \in S_h$) connected to the hub with at least one weekly flight.[4] The equivalent number of spokes is defined as the simple count of spoke airports, each weighted by the routing factor passengers would experience when travelling from the spoke to the final destination via the hub. This variable considers both the hub airport network scope (cardinality of S_h) and its potential to serve a given long-haul route due to its geographical positioning (routing factor effect). A major shortcoming of this measure is that it assigns the same weight to each spoke regardless of the supply (i.e., the number of flights) connecting the spoke to the hub, ultimately underpinning the potential connectivity delivered by the hub to the final destination. To address this, we refine the measure by weighting each spoke proportionally to the frequency of flights to the hub(f_{sh}). This metric, labeled corrected frequency, prioritizes actual connectivity over theoretical connectivity, ultimately better reflecting the potential connectivity offered by the hub for flights to the final destination. Despite these advantages, the corrected frequency still remains a supply-based measure (based on flight schedule) and does not account for actual market size—that is, the number of passengers who could be attracted in the case of the opening of the new long-haul route, given the hub feeding structure. To overcome this limitation, we introduce a third measure: the corrected demand. This measure, in addition to the routing factor, also considers the number

[4]While the concept of spoke was traditionally intended from a local perspective, new Middle Eastern and Turkish hubs rely heavily on long-haul-to-long-haul connectivity. Accordingly, we do not consider an explicit distance threshold to define the set of spoke airports.

of passengers currently traveling from spoke s to final destination d via one-stop itineraries ($\hat{d}_{sd}$) as retrieved from OAG Traffic Analyser. These passengers represent the potential demand that could be attracted in case a new route from hub h to final destination d enters the market. The new itinerary would be much more attractive the higher the market size and the lower the routing factor. Ultimately, the corrected demand captures the size of the feeder base for a given long-haul route, given the airline network.

This set of variables captures the key factors airlines consider when evaluating the launch of a new route. A key advantage of this approach is that these dimensions can be measured not only for existing routes but also for potential ones. More importantly, all the variables presented are publicly available (i.e., externally accessible commercial and open data sources), making them practical tools for investigating network-planning and route-expansion decisions.[5] As discussed, from the broad set of variables presented, we retained for the econometric model only the most relevant—those that significantly enhance the model's predictive power without introducing multicollinearity issues. The final set of explanatory variables used in the empirical model is detailed in Section 4.1.

3.3. Methods

To investigate the effects of various factors on route-entry decisions, we employed a logistic regression model. The model systematically compares the characteristics of newly launched routes with those of potential but unserved routes, allowing the estimation of the probability of a route being opened based on a set of predictors. The estimated coefficients quantify the relationship between each factor and the likelihood of a route entry, where positive coefficients indicate a higher probability of a route being opened and negative coefficients suggest a lower probability. A key advantage of using logistic regression in this context lies in its ability to provide interpretable coefficients, offering clear insights into the contribution of each predictor in shaping route-entry decisions. This straightforward interpretability can facilitate the understanding of the complex interplay of factors—such as demand potential, geographic considerations, network effects, and competitive dynamics—that underpin airline network development strategies. An alternative approach would be to rely on advanced machine-learning classification models, such as random forests. We tested such models and observed only marginal improvements in predictive performance relative to logistic regression, at the cost of significantly lower interpretability. This limited gain is likely due to the relatively small number of positive instances—reflecting the sporadic nature of long-haul route entries—and the constrained availability of high-dimensional explanatory variables from non-proprietary data sources. Therefore, given the objectives of this study, we focus on the results obtained from the logistic regression model.

We divided the data sample described in Section 3.1 into two parts: observations from 2014 to 2017, used for model calibration (training), and observations from 2018 to 2019, used as a testing bed to evaluate model performance. Since new long-haul route entries are relatively rare events, the dataset was highly imbalanced, with the number of potential routes far exceeding the number of actual route openings (48×). Such an imbalance has proven to negatively impact logistic regression, which tends to favor the majority class (Lai et al. 2021, Megahed et al. 2021). To cope with this issue, we trained the model on a balanced dataset, while testing was conducted on the original imbalanced dataset to assess

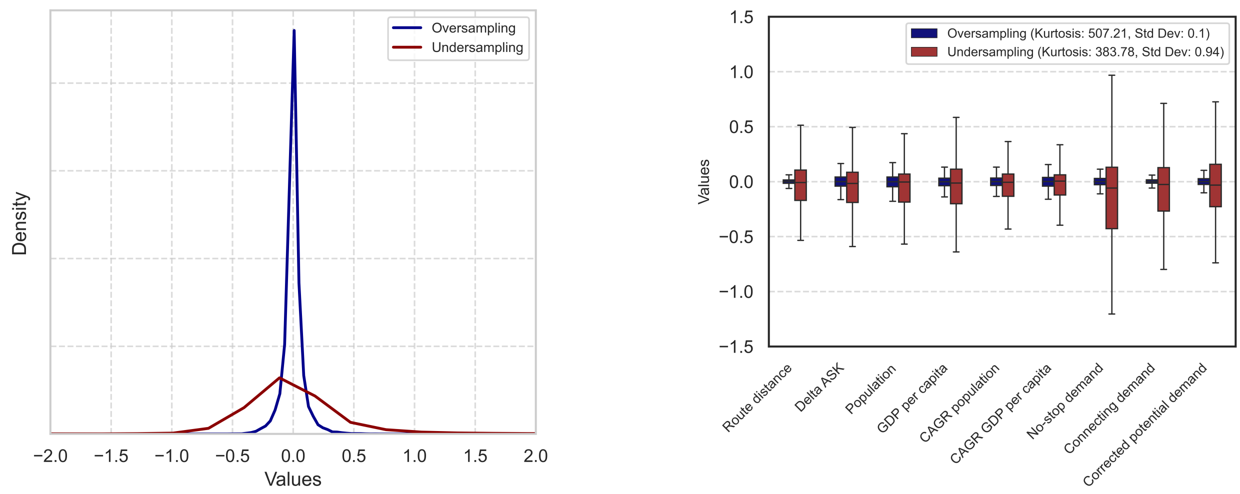
[5]To mitigate potential reverse-causality issues, we used one-year lagged values for all time-dependent variables. Beyond the econometric benefits, this approach better reflects the information available to airlines during their planning process.

its generalizability and performance in real-world conditions. The balanced dataset for training was constructed by applying two common resampling techniques: oversampling and undersampling. Oversampling involves artificially increasing the number of instances in the minority class by duplicating existing observations. This technique ensures that the minority class is adequately represented, providing the model with more opportunities to learn from the underrepresented class (Wang et al. 2018, Ustyannie and Suprapto 2020). On the other hand, undersampling reduces the number of instances in the majority class by randomly discarding observations, thus obtaining a more balanced dataset. While this approach mitigates bias toward the majority class, it risks losing potentially useful data and may lead to a less robust model (Cartus et al. 2020, Kubus 2020). Both methods modify the original data distribution by either increasing the number of samples from the minority class or reducing those from the majority class. These adjustments aim to mitigate class imbalance, enhancing the model's ability to classify (Oommen et al. 2011, Yap et al. 2014). While both approaches are considered valid, there is no strong consensus on which one yields superior performance, since their effectiveness is typically context-dependent. Accordingly, the choice between oversampling and undersampling varies based on the dataset characteristics and study objectives (Chawla et al. 2002, He and Garcia 2009).

To determine which resampling technique would yield the most stable and reliable results for our case, we repeated the logistic regression 100 times, randomly selecting the sample with both techniques. The optimal approach was then determined by evaluating two aspects: coefficient stability and forecasting performance. The former involves assessing coefficient stability, defined here as the robustness of coefficient estimates to repeated random resampling of the training data. Specifically, stability is evaluated by examining how consistently each variable's coefficient remains close to its mean value across multiple model estimations, with lower dispersion indicating greater robustness to resampling. Figure 2a presents the distribution of coefficients around their mean under the two resampling methods, while Figure 2b compares the coefficient distributions across variables for both techniques. Only statistically significant coefficient estimates were included in the analysis. To ensure comparability within the same graph, the coefficients were transformed using a tailored formulation, quantifying their relative deviation from the mean across multiple iterations.[6] This transformation allows deviations to be expressed on a comparable scale across variables with different magnitudes, facilitating the assessment of robustness to resampling strategy. The shape of the curve in Figure 2a shows whether the coefficients cluster around their mean or exhibit high variability. Overall, coefficients obtained through oversampling exhibit lower variability. This finding is also supported when examining single explanatory variables (Figure 2b). Accordingly, we can conclude that, in our case, oversampling reduces sensitivity to variations across iterations, thereby enhancing model robustness.

A second aspect to consider when investigating the preferred resampling technique is model forecasting performance, which is the model's ability to correctly classify instances in the testing dataset. We evaluated model performance by focusing on three key performance indicators (KPIs): accuracy, precision, and sensitivity. *Accuracy* is the most intuitive classification metric, representing the overall proportion of correctly classified instances—both positive and negative—out of all predictions. *Precision*, on the other hand, measures the proportion of positive predictions—that is, routes iden-

[6]For each random resampling of the training data i, we compute coefficient deviation x_i as $\frac{\beta_i-\bar{\beta}}{\bar{\beta}}$, where β_i denotes the coefficient estimate in resampling i and $\bar{\beta}$ is the mean coefficient across all iterations.



(a) All variables

(b) By variable

Figure 2: Coefficient deviation (stability) from their mean over 100 iterations (oversampling vs. undersampling).

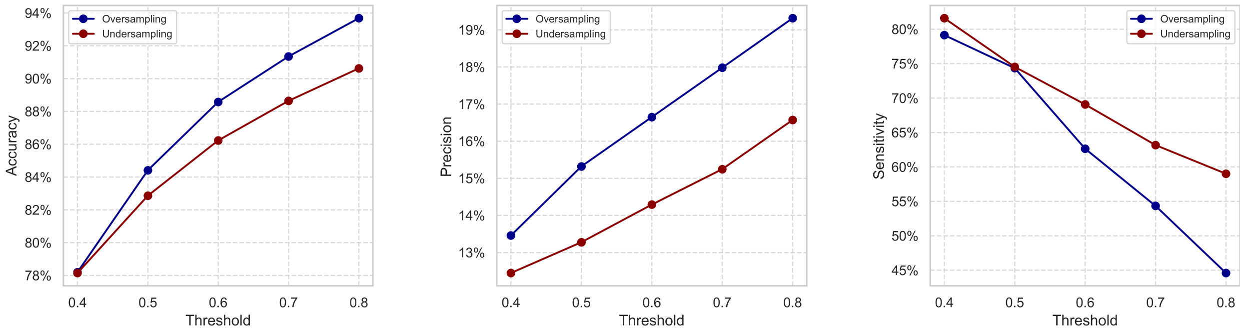
tified by the model as new—that are actually correct and have in fact been opened by the airline.[7] Precision provides insights into the accuracy of the model's predictions when it forecasts a promising new route. Lastly, *sensitivity*—also known as recall—measures the proportion of actual positive instances—namely, new routes—that are correctly identified by the model.[8] In our analysis, sensitivity is particularly important, as it reflects how effectively the model detects new routes that an airline has opened.

Figure 3 shows the three KPIs for the different resampling techniques and the various predicted probability thresholds used to classify routes as promising or not promising. In terms of accuracy, both resampling techniques performed well, ranging from 0.78 to 0.94 depending on the chosen threshold. However, oversampling consistently outperformed undersampling. Similarly, oversampling ensured a precision that was consistently 1%-2% higher than that of undersampling across all threshold levels. Sensitivity, on the other hand, showed the most heterogeneous pattern among the KPIs. At lower thresholds (0.4 and 0.5), oversampling and undersampling showed similar performance. However, as the threshold became more stringent, undersampling demonstrated superior performance.

In summary, our evaluation of resampling techniques based on coefficient stability and forecasting performance suggests that, in our context, oversampling is the preferred approach. It consistently yielded more stable coefficient estimates across different random resampling iterations, indicating lower variability and greater model robustness. Additionally, while both techniques demonstrated good forecasting performance, oversampling achieved higher accuracy and precision. On the other hand, the results for sensitivity were more favorable to undersampling. Overall, these findings indicate that oversampling would be the more reliable and effective balancing technique for our analysis, as it enhances model consistency while improving the precision of route predictions. Accordingly, in

[7] Precision is defined as the proportion of true-positive predictions (TP; i.e., routes predicted by the model as promising that have actually been opened) out of all positive predictions made by the model, including both correct (TP) and incorrect (false-positive, FP) predictions. Formally, it is computed as $\frac{TP}{TP+FP}$.

[8] Sensitivity is defined as the proportion of true-positive predictions (i.e., routes classified as promising and opened by the airline) out of the total number of routes opened by the airline (TP + FN). Formally, sensitivity is computed as $\frac{TP}{TP+FN}$.



(a) Accuracy **(b)** Precision **(c)** Sensitivity

Figure 3: Forecasting performance KPIs (oversampling vs. undersampling).

the following section, we discuss the empirical results obtained using oversampling as a resampling technique.

4. Results

4.1. Empirical results

In this section, we discuss the empirical results of the classification model applied to the comprehensive dataset of new route openings and potential route candidates detailed above. We develop carrier-specific formulations, offering a detailed examination of the coefficients influencing each carrier's network expansion decisions. Additionally, we compare these results with those obtained from a pooled logit model that does not account for carrier-specific segmentation.

For each carrier in our sample, we performed 100 oversampling iterations following the oversampling procedure detailed in Section 3.3. In each iteration, the logit model was fitted using the instances within the training dataset belonging to the selected airline and balanced through oversampling, providing a carrier-specific framework for network-expansion modeling. The model captures the effect of key drivers on airline network expansion strategies and reflects the importance of each variable within the airline decision-making process. To obtain single coefficient estimates for each carrier, we aggregate the results across the different iterations. Specifically, coefficient magnitude is computed as the average across all iterations, while the statistical significance of the estimated coefficients is assessed using the z-score, calculated as the ratio of the mean coefficient to its standard error.

Figure 4 reports the coefficients of the carrier-specific logit models as well as those of the pooled specification. To facilitate interpretation and comparison, coefficients are standardized and displayed using a heatmap.[9] Cell color indicates the sign of the average coefficient, with green denoting a positive coefficient and red a negative one, while color intensity reflects the relative magnitude of the effect across variables for the same airline. More saturated colors correspond to factors with a stronger association with the likelihood of route opening, whereas lighter or near-white cells indicate variables with a weak or statistically non-significant contribution. This visualization provides a concise overview of the direction and relative importance of key factors across airlines.

[9]The standardization procedure follows the methods outlined in Adler and Hashai (2005), Menard (2011), and Karlson (2015), which scale coefficients based on the variance of the predictors, thereby ensuring comparability across variables and facilitating the interpretation of their relative importance within each airline's decision-making process.

The heterogeneous results, both in terms of coefficient magnitude and direction, highlight the complex interplay of factors shaping the network expansion strategies of different carriers. In contrast, the soft colors of the pooled model reflect coefficient attenuation due to aggregation across heterogeneous carriers, suggesting that a generalized approach fails to adequately capture the complexities of carrier-specific expansion dynamics. Looking at individual variables, the most influential factor across all carriers is connecting demand, which consistently exhibits a strong and positive effect on the likelihood of entering a route. This result underscores the importance of serving a market with a nonstop connection, potentially attracting passengers who are currently traveling on connecting itineraries. Nonstop demand and the strength of potential demand from the spokes also emerge as critical drivers, although their impact varies across carriers. Considering nonstop demand, some carriers, such as Delta and Air Canada, tend to favor entry into routes already served by nonstop services. In contrast, other airlines, including Emirates and Turkish Airlines, exhibit a tendency to avoid opening new routes in markets already served by nonstop services. The potential feeder base and the positioning of the hub airport to effectively serve a given destination, considering the existing network, are determining factors for many carriers. This is the case for Gulf carriers such as Emirates and Qatar Airways, as well as Turkish Airlines, Korean Air, and LATAM Airlines, which demonstrate a strong reliance on connecting traffic to support new route openings. Consistent with prior research, distance exhibits a negative effect on route openings across all airlines. Socioeconomic variables do not display a homogeneous role in determining route opening. On the one hand, some carriers, such as Ethiopian Airlines and British Airways, place great emphasis on current market conditions, prioritizing socio-economic factors when making expansion decisions. On the other hand, airlines such as KLM and Air France focus on future growth potential, aligning their strategies with long-term-oriented demographic and economic projections. Lastly, some airlines prove to be strongly dependent on fleet and capacity considerations (ASK variable) when determining their route network expansion. This is particularly evident for KLM, which tends to plan network adjustments based on the growth potential (in terms of capacity) available for each type of destination. Overall, this fragmented landscape highlights the heterogeneity in airlines' approaches, further demonstrating the benefits of carrier-specific modeling over a generalized framework.

We evaluate model performance using a customized out-of-sample validation approach applied to the unbalanced testing dataset, considering different KPIs, such as accuracy, precision, and sensitivity, as discussed in Section 3.3. To complement these metrics, we introduce the *screening ratio*, a tailored KPI designed to assess the model's ability to narrow down the set of potential routes and identify the most promising ones. It is defined as the proportion of routes identified by the model as promising relative to the total number of candidates. This metric serves as a proxy for the model's effectiveness in selecting a subset of candidate routes to be considered for further investigation using network-planning optimization algorithms. Ultimately, it quantifies the potential reduction in the complexity of the network-planning algorithm when fed by the results of the proposed model.

Table 4 reports the out-of-sample performance of the carrier-specific models and the pooled logit model that does not account for carrier-specific variations. As expected, airlines exhibiting lower model goodness-of-fit and accuracy are also those with the fewest new routes introduced during the training period (i.e., from 2014 to 2017). This limited expansion reduces the amount of historical data used for model calibration, thereby constraining the model's ability to achieve high performance,

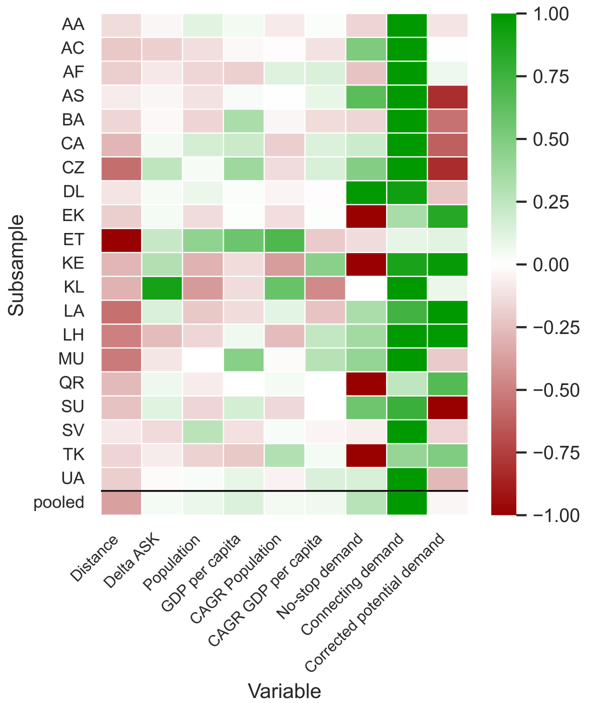


Figure 4: Model results (heatmap with standardized coefficients).

Carrier	Pseudo-R^2	Accuracy	Precision	Sensitivity	Screening ratio
AA	0.69	94.6%	29.4%	78.7%	6.9%
AC	0.42	83.5%	10.8%	84.9%	18.1%
AF	0.31	86.3%	6.4%	35.0%	13.0%
AS	0.70	93.8%	86.6%	93.5%	32.2%
BA	0.32	74.6%	7.3%	90.0%	27.2%
CA	0.27	82.2%	11.0%	73.3%	19.2%
CZ	0.37	82.8%	14.6%	65.0%	18.5%
DL	0.79	97.0%	52.0%	86.7%	5.3%
EK	0.19	81.2%	5.4%	72.7%	19.5%
ET	0.31	71.2%	4.7%	83.3%	29.9%
KE	0.66	86.3%	4.3%	100.0%	14.3%
KL	0.14	68.6%	3.3%	90.9%	32.4%
LA	0.57	88.5%	8.6%	84.8%	12.3%
LH	0.43	87.5%	4.9%	54.5%	12.6%
MU	0.23	78.4%	7.2%	77.3%	22.7%
QR	0.22	79.9%	8.8%	45.2%	19.8%
SU	0.81	96.7%	14.7%	50.0%	3.3%
SV	0.35	81.9%	5.2%	88.0%	18.9%
TK	0.47	84.1%	10.1%	58.3%	16.4%
UA	0.61	90.5%	12.6%	82.6%	10.5%
Pooled	0.29	84.5%	9.2%	68.0%	16.3%

Table 4: Model KPIs considering a threshold of 0.5 to identify promising routes.

particularly in comparison to carriers with more extensive route development. While the pooled model, which aggregates all airlines into a single framework, achieves a relatively high classification accuracy (84.5%), its goodness-of-fit and precision remain limited, with a pseudo-R^2 of approximately 0.29 and a precision of 9.2%. This contrast highlights the inability of a one-size-fits-all specification to capture the heterogeneity underlying individual airlines' route-entry decisions, reinforcing the benefits of carrier-specific modeling approaches.

To further explore model performance, the models were tested at varying probability thresholds (ranging from 0.4 to 0.8), rather than the canonical 0.5 threshold, to define promising routes. By varying the threshold, we simulated different risk tolerance levels that airlines might have had when making decisions about new route openings. Lower thresholds (e.g., 0.4) may lead to more aggressive predictions—resulting in a higher number of promising routes identified—but possibly at the cost of precision. Conversely, higher thresholds (e.g., 0.8) might lead to fewer predictions, potentially increasing precision but sacrificing sensitivity. This trade-off is clearly depicted in Figure 5, which presents a boxplot of three KPIs—sensitivity, precision, and screening ratio—across different thresholds. As the threshold increases, precision tends to improve (fewer false positives), but this may come at the cost of sensitivity—that is, a decrease in true positives. Conversely, lowering the threshold may improve accuracy but lead to a reduction in precision as more false positives are generated. By analyzing these results, we gain valuable insights into the model's effectiveness in identifying new routes under different decision making scenarios. This analysis not only highlights the model's sensitivity to threshold variations but also helps determine the optimal threshold for identifying new routes while balancing the competing priorities of sensitivity, precision, and screening ratio. A detailed breakdown of confusion-matrix elements and KPIs for each carrier across different probability thresholds to identify

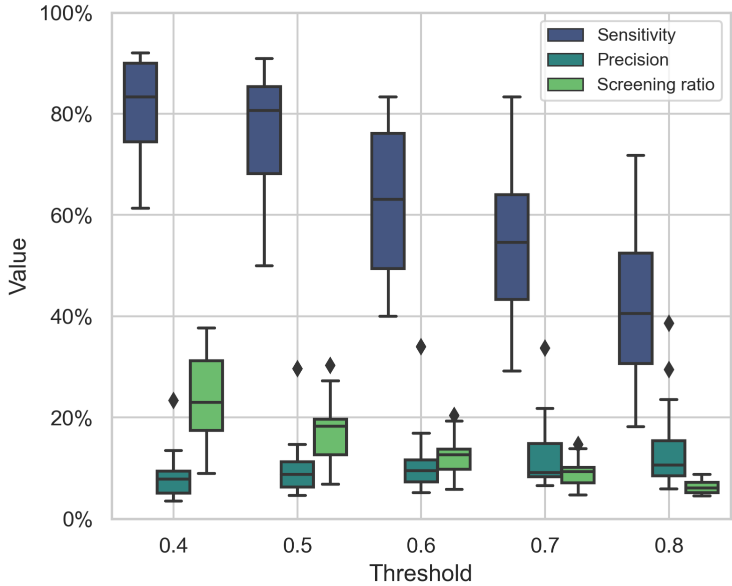


Figure 5: KPIs across different threshold levels.

promising routes is reported in Appendix A (Table A.1).

4.2. Airline case study

This section demonstrates how the proposed classification model can be used to identify promising routes through case studies of three major airlines: British Airways (BA), Qatar Airways (QR), and Turkish Airlines (TK). To assess the effectiveness of the model, we analyze the network expansion opportunities suggested by the model—defined as routes with high predicted probabilities based on 2019 data. These predictions are then compared to the actual new routes launched by airlines between 2019 and 2024. Figure 6 illustrates the routes classified as promising by the model, along with long-haul destinations already served by the airlines in 2019.

British Airways, the flag carrier of the United Kingdom, operates an extensive global network that centers on its main hub at London Heathrow (LHR). In 2019, the airline maintained a strong transatlantic presence, with flights to major North American destinations, such as New York (JFK), Boston (BOS), Toronto (YYZ), and Chicago (ORD). Beyond North America, British Airways also had well-established connections to the Middle East, South Asia, and parts of Africa. Consistent with these network characteristics and past entry dynamics, the classification model identifies several destinations as promising candidates for further network expansion in fast-growing regions (Figure 6a). Southeast Asia, in particular, stands out as a key area for future expansion consistent with growth dynamics observed in socioeconomic drivers. The model also indicates potential new destinations in Central Asia and North America. Notably, six of the destinations classified as promising by the model had been added to the British Airways network by 2024: Islamabad (ISB) and Malé (MLE) in Asia, Paphos (PFO) in Cyprus, Hamilton (BDA) in the Bermuda Islands, and Cincinnati (CVG) and Pittsburgh (PIT) in the United States.

A similar pattern emerges for Qatar Airways. The rapidly growing Middle Eastern carrier established itself during the previous decade as a leading global airline, with a well-developed network linking the Middle East to Europe, Asia, Africa, and the Americas. By 2019, the airline's operations had become heavily concentrated around its hub in Doha (DOH), serving as a central gateway connecting long-haul markets. The model indicates network expansion opportunities for the airline in some destinations in Asia and Africa, as well as strengthening its presence in secondary European and North American destinations (see Figure 6b). These destinations align with the airline's long-haul-to-long-haul hub strategy, which leverages connecting demand across continents. Since 2019, nine of the 22 promising routes have been opened as new nonstop destinations by the airline: Lisbon (LIS), Hamburg (HAM), and Lyon (LYS) in Europe, Seattle (SEA) and Pittsburgh (PIT) in North America, Accra (ACC) in Africa, Cebu (CEB) and Osaka (KIX) in Asia, and Brisbane (BNE) in Oceania.

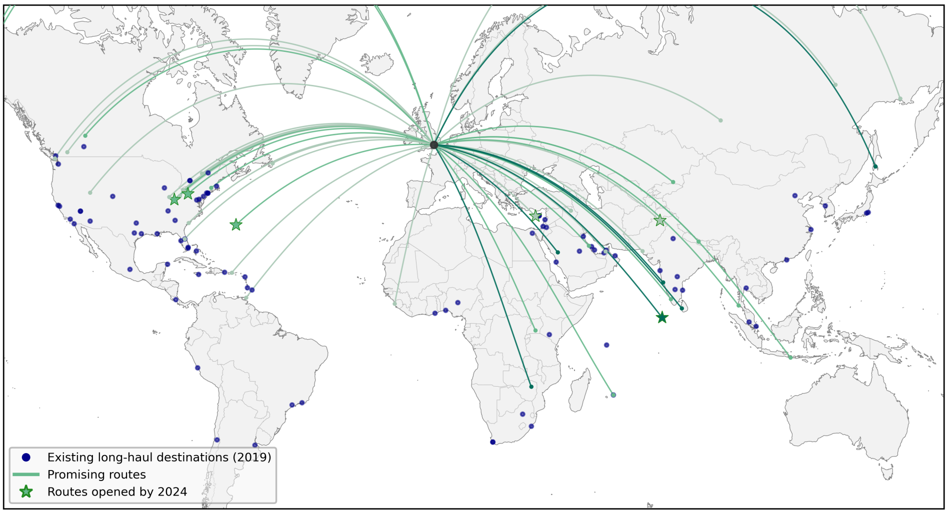
The last carrier analyzed is Turkish Airlines (Figure 6c). This airline operates one of the largest and most diverse route networks globally, connecting long-haul destinations across Asia, Africa, and North America to its hub in Istanbul (IST), which plays a crucial role as a strategic transit point between East and West. The model suggests that Turkish Airlines could further expand its network by strengthening its presence in Central Asia—a region with a growing population—as well as in Africa, particularly South Africa and along the west coast. Notably, 10 out of 27 destinations identified as promising by the model were included in the airline's network between 2019 and 2024: Seattle (SEA), Vancouver (YVR), Marrakesh (RAK), N'Djamena (NDJ), Abidjan (ABJ), Luanda (LAD), Osaka (KIX), Mexico City (MEX), Denpasar (DPS), and Colombo (CMB)."

Overall, applying the classification model to the three selected airlines demonstrated that the model can effectively identify promising routes for consideration by airlines for network development.

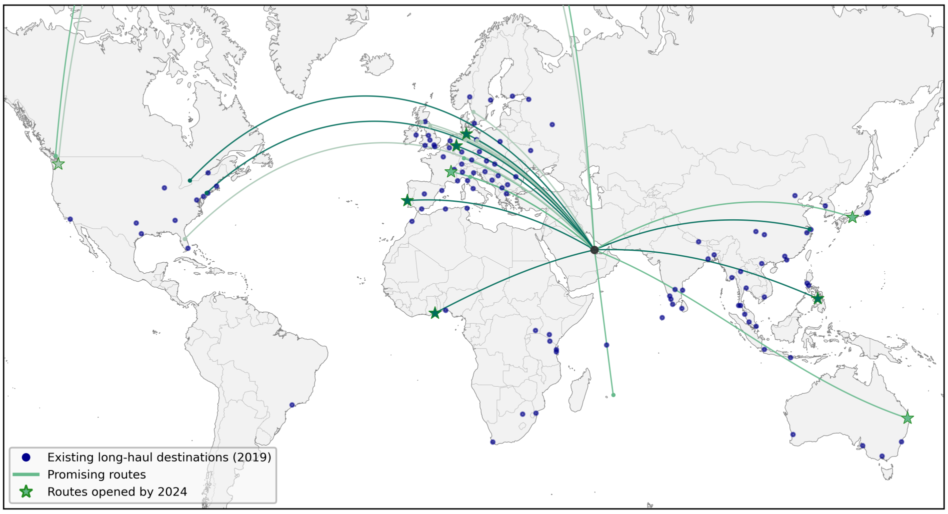
5. Conclusion

This study aimed to empirically investigate the key factors that shape airline decisions to enter new long-haul markets. To this end, we developed a tailored classification model applied to a global dataset spanning six years (2014–2019) to investigate promising routes based on a refined set of features derived from publicly available data. These features capture key dimensions such as market potential, competitive intensity, and network connectivity, enabling a structured and data-driven exploration of factors shaping airline route-development decisions. Faced with a highly unbalanced dataset, we compared different resampling techniques and found that, in our context, oversampling yielded more stable and robust results than undersampling.

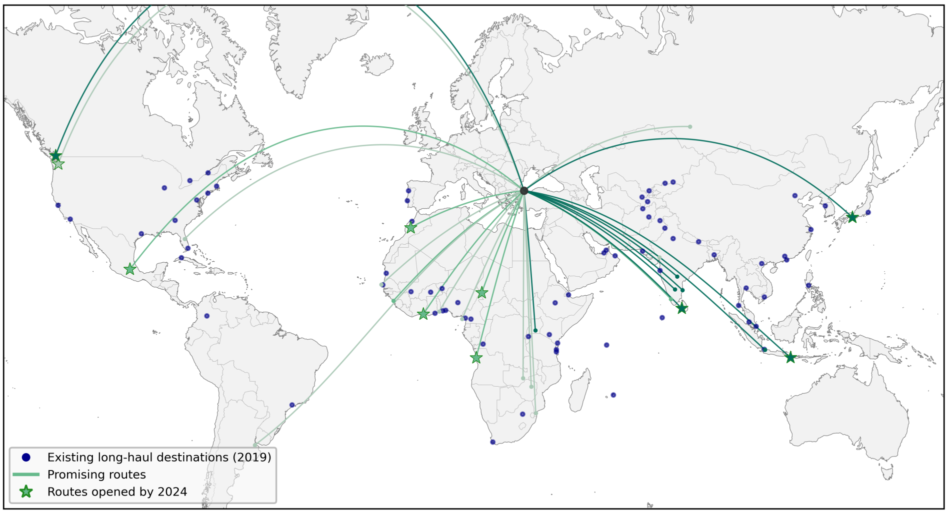
The results empirically highlight substantial heterogeneity in route selection patterns across airlines. Carrier-specific models clearly outperform the pooled specification. Among the key drivers, connecting demand emerges as a critical factor considered by airlines when assessing the possibility of launching a new route. Some airlines appear more inclined to prioritize markets with strong socioeconomic growth potential, while others place greater emphasis on fleet availability and network connectivity. This diversity reinforces the need for customized predictive tools for individual carrier profiles. The practical applicability of the proposed model was demonstrated through case studies involving three major airlines. In each case, the model demonstrated its effectiveness in identifying potential routes that would subsequently be launched, providing evidence of its applicability to real-world network-planning decisions. Overall, the model proved effective in narrowing down a set of promising routes and can serve as a pre-screening tool, especially when integrated with airline proprietary data.



(a) British Airways (BA).



(b) Qatar Airways (QR).



(c) Turkish Airlines (TK).

Figure 6: Promising routes for three selected airlines, as identified by the classification model applied to 2019 data. Darker lines indicate higher predicted probabilities.

Despite the model's practical applicability and the insights it provides, predicting new route openings with publicly available data remains an exceptionally challenging task for several reasons. First, route entry is relatively rare. Accordingly, when examining a large period, the actual number of new long-haul routes launched is relatively small, thus constraining the sample size and, in turn, the model's predictive power. Second, although the model incorporates a comprehensive set of factors, it is difficult to gather determinant variables with global coverage and sufficient granularity. Examples include data concerning events, public holidays, and tourism intensity—factors that often influence airline decisions but are not consistently available on a global scale. The scarcity of detailed and high-quality data and the relatively small sample size also reduce the applicability of machine learning techniques. When testing such methods, we obtained poorer results than with the classification approach adopted in this study. Future research could benefit from the availability of richer data sources and big data, including detailed tourism indicators, pricing information, and event-level data, which could further enhance predictive performance and potentially unlock the full potential of machine-learning techniques in this domain.

Acknowledgments

The authors sincerely thank Augustin Lesgourgues of Air France for the fruitful discussion on the preliminary version of the paper and for his valuable insights into airlines' decision-making process regarding new route development.

References

Abdelghany A, Guzhva VS (2010) Analyzing airlines market service using panel data. *Journal of Air Transport Management* 16(1):20–25.

Abdelghany A, Guzhva VS (2022) Exploratory analysis of air travel demand stimulation in first-time served markets. *Journal of Air Transport Management* 98:102162.

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.

Alumur SA, Campbell JF, Contreras I, Kara BY, Marianov V, O'Kelly ME (2021) Perspectives on modeling hub location problems. *European Journal of Operational Research* 291(1):1–17.

Aydemir R (2012) Threat of market entry and low cost carrier competition. *Journal of Air Transport Management* 23:59–62.

Bannò M, Redondi R (2014) Air connectivity and foreign direct investments: economic effects of the introduction of new routes. *European Transport Research Review* 6:355–363.

Baykasoğlu A, Dudaklı N, Subulan K, Taşan AS (2022) An integrated fleet planning model with empty vehicle repositioning for an intermodal transportation system. *Operational Research* 22(3):2063–2098.

Birolini S, Jacquillat A, Cattaneo M, Antunes AP (2021) Airline network planning: Mixed-integer non-convex optimization with demand–supply interactions. *Transportation Research Part B: Methodological* 154:100–124.

Boguslaski C, Ito H, Lee D (2004) Entry patterns in the southwest airlines route system. *Review of Industrial Organization* 25:317–350.

Calzada J, Fageda X (2019) Route expansion in the european air transport market. *Regional Studies* 53(8):1149–1160.

Carmona-Benítez RB, Nieto MR, Miranda D (2017) An econometric dynamic model to estimate passenger demand for air transport industry. *Transportation Research Procedia* 25:17–29.

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.

Cartus AR, Bodnar LM, Naimi AI (2020) The impact of undersampling on the predictive performance of logistic regression and machine learning algorithms: a simulation study. *Epidemiology* 31(5):e42–e44.

Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP (2002) Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research* 16:321–357.

Cheung TK, Wong CW, Lei Z (2022) Assessment of hub airports' connectivity and self-connection potentials. *Transport Policy* 127:250–259.

Dresner M, Eroglu C, Hofer C, Mendez F, Tan K (2015) The impact of gulf carrier competition on us airlines. *Transportation Research Part A: Policy and Practice* 79:31–41.

Fu X, Lei Z, Wang K, Yan J (2015) Low cost carrier competition and route entry in an emerging but regulated aviation market–the case of china. *Transportation Research Part A: Policy and Practice* 79:3–16.

Gaggero AA, Piazza G (2021) Multilayer networks and route entry into the airline industry: Evidence from the us domestic market. *Research in Transportation Economics* 90:101044.

Geursen IL, Santos BF, Yorke-Smith N (2023) Fleet planning under demand and fuel price uncertainty using actor–critic reinforcement learning. *Journal of Air Transport Management* 109:102397.

Halpern N, Graham A (2015) Airport route development: A survey of current practice. *Tourism Management* 46:213–221.

Hanson D, Delibasi TT, Gatti M, Cohen S (2022) How do changes in economic activity affect air passenger traffic? the use of state-dependent income elasticities to improve aviation forecasts. *Journal of Air Transport Management* 98:102147.

Hausladen I, Schosser M (2020) Towards a maturity model for big data analytics in airline network planning. *Journal of Air Transport Management* 82:101721.

He H, Garcia EA (2009) Learning from imbalanced data. *IEEE Transactions on knowledge and data engineering* 21(9):1263–1284.

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

Karlson KB (2015) Another look at the method of y-standardization in logit and probit models. *The Journal of Mathematical Sociology* 39(1):29–38.

Kim A, Hansen M (2015) Some insights into a sequential resource allocation mechanism for en route air traffic management. *Transportation Research Part B: Methodological* 79:1–15.

Kölker K, Lütjens K (2015) Using genetic algorithms to solve large-scale airline network planning problems. *Transportation research procedia* 10:900–909.

Kubus M (2020) Evaluation of resampling methods in the class unbalance problem. *Econometrics. Ekonometria. Advances in Applied Data Analytics* 24(1):39–50.

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):1–15.

Lai SBS, Shahri N, Mohamad MB, Rahman H, Rambli AB (2021) Comparing the performance of adaboost, xgboost, and logistic regression for imbalanced data. *Mathematics and Statistics* 9(3):379–385.

Lee SY, Yoo KE, Park Y (2014) A continuous connectivity model for evaluation of hub-and-spoke operations. *Transportmetrica A: Transport Science* 10(10):894–916.

Megahed FM, Chen YJ, Megahed A, Ong Y, Altman N, Krzywinski M (2021) The class imbalance problem.

Menard S (2011) Standards for standardized logistic regression coefficients. *Social Forces* 89(4):1409–1428.

Mohammadi M, Jula P, Tavakkoli-Moghaddam R (2019) Reliable single-allocation hub location problem with disruptions. *Transportation Research Part E: Logistics and Transportation Review* 123:90–120.

Mohri SS, Nasrollahi M, Pirayesh A, Mohammadi M (2022) An integrated global airline hub network design with fleet planning. *Computers & Industrial Engineering* 164:107883.

Morrison SA, Winston C (1990) The dynamics of airline pricing and competition. *The American Economic Review* 80(2):389–393.

Oliveira AV (2008) An empirical model of low-cost carrier entry. *Transportation Research Part A: Policy and Practice* 42(4):673–695.

Oommen T, Baise LG, Vogel RM (2011) Sampling bias and class imbalance in maximum-likelihood logistic regression. *Mathematical Geosciences* 43:99–120.

Santana M, De La Vega J, Morabito R, Pureza V (2023) The aircraft recovery problem: A systematic literature review. *EURO Journal on Transportation and Logistics* 100117.

Schosser M, Schosser M (2020) Status quo of strategic network planning in airlines. *Big Data to Improve Strategic Network Planning in Airlines* 161–194.

Sharma A, Jakhar SK, Vlachos I, Kumar S (2024) Advances in hub location problems: a literature review and research agenda. *International Journal of Productivity and Performance Management* .

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.

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: Methodological* 28(2):103–121.

Teoh LE, Khoo HL (2016) Fleet planning decision-making: Two-stage optimization with slot purchase. *Journal of Optimization* 2016(1):8089794.

Ustyannie W, Suprapto S (2020) Oversampling method to handling imbalanced datasets problem in binary logistic regression algorithm. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)* 14(1):1–10.

Wang H, Zhu R, Ma P (2018) Optimal subsampling for large sample logistic regression. *Journal of the American Statistical Association* 113(522):829–844.

Wang K, Tsui KWH, Liang L, Fu X (2017) Entry patterns of low-cost carriers in hong kong and implications to the regional market. *Journal of Air Transport Management* 64:101–112.

Wang X, Peng J, Tang J, Lu Q, Li X (2022) Investigating the impact of adding new airline routes on air transportation resilience in china. *Transport Policy* 125:79–95.

Wen X, Chung SH, Ma HL, Khan WA (2024) Airline crew scheduling with sustainability enhancement by data analytics under circular economy. *Annals of Operations Research* 342(1):959–985.

Wong CW, Cheung TKY, Zhang A (2023) A connectivity-based methodology for new air route identification. *Transportation Research Part A: Policy and Practice* 173:103715.

Wu J, Zhang Pw, Wang Y, Shi JJ (2022) Integrated aviation model and metaheuristic algorithm for hub-and-spoke network design and airline fleet planning. *Transportation Research Part E: Logistics and Transportation Review* 164:102755.

Xu Y, Adler N, Wandelt S, Sun X (2024) Competitive integrated airline schedule design and fleet assignment. *European Journal of Operational Research* 314(1):32–50.

Yap BW, Rani KA, Rahman HAA, Fong S, Khairudin Z, Abdullah NN (2014) An application of oversampling, undersampling, bagging and boosting in handling imbalanced datasets. *Proceedings of the first international conference on advanced data and information engineering (DaEng-2013)*, 13–22 (Springer).

Zhang Y, Wang K, Fu X (2017) Air transport services in regional australia: Demand pattern, frequency choice and airport entry. *Transportation Research Part A: Policy and Practice* 103:472–489.

Zografos KG, Salouras Y, Madas MA (2012) Dealing with the efficient allocation of scarce resources at congested airports. *Transportation Research Part C: Emerging Technologies* 21(1):244–256.

Zou L, Yu C (2020) The evolving market entry strategy: A comparative study of southwest and jetblue. *Transportation Research Part A: Policy and Practice* 132:682–695.

Appendix A. Threshold analysis

Table A.1: Confusion-matrix elements and KPIs considering different probability thresholds to classify routes as promising.

Threshold	Metric	AA	AC	AF	AS	BA	CA	CZ	DL	EK	ET	KE	KL	LA	LH	MU	QR	SU	SV	TK	UA	Pooled
	True Positive	38	67	9	58	18	25	30	53	8	15	6	10	29	14	38	19	5	23	15	38	514
0.4	True Negative	1,649	2,441	680	134	572	731	731	1,778	523	669	794	494	2,273	1,606	1,345	547	993	1,410	657	2,478	21,191
	False Positive	127	651	146	12	310	283	193	56	226	387	164	433	343	319	677	226	35	794	162	350	6,696
	False Negative	9	6	11	4	2	5	10	7	3	3	0	1	4	8	6	12	5	2	9	8	115
	Accuracy	93%	79%	81%	92%	65%	72%	79%	97%	70%	64%	83%	54%	87%	83%	67%	70%	96%	64%	80%	88%	76%
	Precision	23%	9%	6%	83%	5%	8%	13%	49%	3%	4%	4%	2%	8%	4%	5%	8%	13%	3%	8%	10%	7%
	Sensitivity	81%	92%	45%	94%	90%	83%	75%	88%	73%	83%	100%	91%	88%	64%	86%	61%	50%	92%	63%	83%	82%
	Screening Ratio	9%	23%	18%	34%	36%	30%	23%	6%	31%	37%	18%	47%	14%	17%	35%	30%	4%	37%	21%	14%	25%
	True Positive	37	62	7	58	18	22	26	52	8	15	6	10	28	12	34	14	5	22	14	38	428
0.5	True Negative	1,687	2,580	723	137	655	836	772	1,786	609	750	826	633	2,317	1,692	1,586	628	999	1,804	695	2,564	23,664
	False Positive	89	512	103	9	227	178	152	48	140	306	132	294	299	233	436	145	29	400	124	264	4,223
	False Negative	10	11	13	4	2	8	14	8	3	3	0	1	5	10	10	17	5	3	10	8	201
	Accuracy	95%	83%	86%	94%	75%	82%	83%	97%	81%	71%	86%	69%	89%	88%	78%	80%	97%	82%	84%	91%	84%
	Precision	29%	11%	6%	87%	7%	11%	15%	52%	5%	5%	4%	3%	9%	5%	7%	9%	15%	5%	10%	13%	9%
	Sensitivity	79%	85%	35%	94%	90%	73%	65%	87%	73%	83%	100%	91%	85%	55%	77%	45%	50%	88%	58%	83%	68%
	Screening Ratio	7%	18%	13%	32%	27%	19%	18%	5%	19%	30%	14%	32%	12%	13%	23%	20%	3%	19%	16%	11%	16%
	True Positive	36	53	7	58	15	15	22	51	5	13	5	6	27	10	21	10	4	19	12	38	369
0.6	True Negative	1,706	2,668	754	139	724	892	814	1,790	686	817	850	741	2,361	1,742	1,749	677	1,003	1,964	718	2,628	25,044
	False Positive	70	424	72	7	158	122	110	44	63	239	108	186	255	183	273	96	25	240	101	200	2,843
	False Negative	11	20	13	4	5	15	18	9	6	5	1	5	6	12	23	21	6	6	12	8	260
	Accuracy	96%	86%	90%	95%	82%	87%	87%	97%	91%	77%	89%	80%	90%	90%	86%	85%	97%	89%	87%	93%	89%
	Precision	34%	11%	9%	89%	9%	11%	17%	54%	7%	5%	4%	3%	10%	5%	7%	9%	14%	7%	11%	16%	11%
	Sensitivity	77%	73%	35%	94%	75%	50%	55%	85%	45%	72%	83%	55%	82%	45%	48%	32%	40%	76%	50%	83%	59%
	Screening Ratio	6%	15%	9%	31%	19%	13%	14%	5%	9%	23%	12%	20%	11%	10%	14%	13%	3%	12%	13%	8%	11%
	True Positive	28	35	5	58	11	14	20	50	4	11	4	6	23	10	18	6	4	16	7	37	319
0.7	True Negative	1,720	2,779	771	142	761	934	853	1,796	703	903	870	803	2,394	1,783	1,833	696	1,007	2,067	744	2,678	25,853
	False Positive	56	313	55	4	121	80	71	38	46	153	88	124	222	142	189	77	21	137	75	150	2,034
	False Negative	19	38	15	4	9	16	20	10	7	7	2	5	10	12	26	25	6	9	17	9	310
	Accuracy	96%	89%	92%	96%	86%	91%	91%	97%	93%	85%	91%	86%	91%	92%	90%	87%	97%	93%	89%	94%	92%
	Precision	33%	10%	8%	94%	8%	15%	22%	57%	8%	7%	4%	5%	9%	7%	9%	7%	16%	10%	8%	20%	14%
	Sensitivity	60%	48%	25%	94%	55%	47%	50%	83%	36%	61%	67%	55%	70%	45%	41%	19%	40%	64%	29%	80%	51%
	Screening Ratio	5%	11%	7%	30%	15%	9%	9%	5%	7%	15%	10%	14%	9%	8%	10%	10%	2%	7%	10%	6%	8%
	True Positive	22	30	5	58	8	12	20	50	2	9	4	4	21	7	12	3	2	15	4	33	260
0.8	True Negative	1,740	2,847	787	142	819	975	876	1,798	715	968	882	873	2,432	1,815	1,914	724	1,007	2,107	771	2,710	26,496
	False Positive	36	245	39	4	63	39	48	36	34	88	76	54	184	110	108	49	21	97	48	118	1,391
	False Negative	25	43	15	4	12	18	20	10	9	9	2	7	12	15	32	28	8	10	20	13	369
	Accuracy	97%	91%	94%	96%	92%	95%	93%	98%	94%	91%	92%	93%	93%	94%	93%	90%	97%	95%	92%	95%	94%
	Precision	38%	11%	11%	94%	11%	24%	29%	58%	6%	9%	5%	7%	10%	6%	10%	6%	9%	13%	8%	22%	16%
	Sensitivity	47%	41%	25%	94%	40%	40%	50%	83%	18%	50%	67%	36%	64%	32%	27%	10%	20%	60%	17%	72%	41%
	Screening Ratio	3%	9%	5%	30%	8%	5%	7%	5%	5%	9%	8%	6%	8%	6%	6%	6%	2%	5%	6%	5%	6%