Empirical exploration of air passenger airport choices and airport catchments across a large region

Kaleab Woldeyohannes Yirgu a, Amy M. Kim a,b*

a Department of Civil Engineering, The University of British Columbia, 6250 Applied Science Lane, Vancouver, BC V6T 1Z4 Canada

b Department of Civil and Environmental Engineering, University of Alberta, 116 St NW, Edmonton, AB T6G 2E1 Canada

* Corresponding author

Email addresses: virgu123@mail.ubc.ca (Kaleab W. Yirgu), amykim@civil.ubc.ca (Amy M. Kim)
Abstract

Disparities in air services among airports across North America have deepened considerably over the last two decades and continue to do so, encouraging air passengers to drive long distances to access airports offering better air services. Decades of studies have documented air passengers from an airport’s presumed catchment area – usually defined by administrative or geographic boundaries – “leaking” across these boundaries to start their air trips from larger, more distant airports. Towards abandoning the implicit catchment assumptions or focus on one specific airport previously relied on to study this issue, we create a novel dataset that combines air tickets purchased by passengers from airports throughout a large section of the US Midwest, together with publicly available data, to better understand the service-based drivers of these choices. Key results include the following. First, changes in air service attributes for short- and medium-length routes at an airport will have a larger influence on air passengers’ airport choices, compared with long routes. Second, comparable proportional changes in air service attributes lead to much larger changes in the market shares and thus, catchment areas, of small and medium airports, compared with large hubs. Within these service characteristics, airfare is found to impact passengers’ airport preferences more than flight frequency and nonstop services. Finally, the catchment areas of small- and medium-size airports are strongly influenced by the proximity and characteristics of neighboring airports. Those of large airports are not, and in fact large airports retain very strong market shares across multiple jurisdictions. The study results further our understanding of the influence of air service attributes on long-distance airport choice across a large geographic area served by a diverse set of airports, and implications on multimodal transportation planning within and across regions.

Keywords: Airport catchment, long-distance airport choice, long-distance air passenger “leakage,” airport passenger market areas, air ticket data, US Midwest.
1. Introduction

Airports have largely been studied with the premise that they have geographically or administratively defined passenger catchment areas, while in reality this is much more complex (Ashford and Benchemam, 1987; de Neufville et al., 2013). Large airports with better air services including nonstop flights, more flight options and cheaper fares will often draw passengers from the assumed catchment areas of smaller, less serviced airports in neighboring regions.

The disparities in air services offered at different airports have been further exacerbated by airline mergers, network reorganizations, personnel shortages, and external shocks such as the 9/11 attack, the 2008 financial crisis and most recently, the COVID-19 global pandemic. The phenomenon of air passengers traveling longer ground distances to utilize airports with better air services compared to those of their closest has been termed “leakage” in the literature. To better understand the characteristics and drivers of leakage, researchers and practitioners have collected data on air passengers’ airport choices through intercept, mail and phone surveys (Suzuki, Crum, and Audino, 2003; Blackstone, Buck, and Hakim, 2006; Suzuki, 2007; de Luca, 2012). These surveys and subsequent analysis results include important details about individual passenger characteristics. However, these studies are also limited in geographic scope and thus, do not (and are not meant to) capture passenger airport choices and “leakage” across large areas spanning multiple administrative boundaries (Gupta, Vovsha, and Donnelly, 2008), inherently leading to studies restricted by a-priori assumptions about the geographic bounds of “leakage”. This limitation is partly due to the lack of coordination between neighboring regional planning agencies (Oden and Sciara, 2020; Rahman, Sciara, and Ryerson, 2021), which prohibits passenger data collection over large areas falling under multiple planning jurisdictions.

Towards a more geographically expansive understanding of how passenger airport choices are influenced by the air services offered by their surrounding airports, we create a novel dataset of air tickets sold from a large portion of the US Midwest spanning multiple cities, counties, states, and airports, combined with publicly available data. The dataset, although without passenger-specific information, offers a large geographic coverage unprecedented in previous studies and is used to inform models of airport choice. The results are used to map airport passenger markets based on air service characteristics and ground travel distances, and explore how these markets change with respect to widening disparities in specific air service characteristics among airports.
2. Literature Review

This literature review discusses: 1) changes in the aviation sector driving air service disparities across airports, 2) airport choice studies in the context of interregional passenger “leakage”, and 3) airport catchment definitions and data collection approaches used to study “leakage.”

2.1 Changes Driving Disparities in Air Services within and across Regions

The attractiveness of an airport to an air traveler is largely determined by the types of services provided at that airport and the characteristics of ground access. Following the 1978 US Airline Deregulation Act, airlines quickly reorganized their routes and consolidated air services at selected hubs, leaving smaller airports with less attractive flight options (Kahn, 1988). Kanafani and Abbas (1987) and Kaemmerle (1991) show that, post-Deregulation, a considerable number of air passengers bypassed such smaller airports in favor of more distant hub airports with lower fares and more diverse flight schedules. Deregulation also encouraged new low-cost carriers (LCCs) to enter the market at some major hubs, further drawing passengers (Gillen and Lall, 2004; Graham, 2013; Atallah and Hotle, 2019).

Some significant airline mergers and alliances, and network reorganizations, occurred in the first 15 years of this century (Ryerson and Kim, 2014). Newly merged airlines further consolidated services at key hubs, and reduced services on redundant and less profitable routes typically operated by regional aircraft (Ryerson and Kim, 2013). For instance, major airports in the US northeast and leisure regions like Florida saw air service expansions, while airports in less populous regions such as the Rust Belt and the Inter-mountain West saw significant reductions (Fuellhart et al. 2016). Such changes have led to further shifts in air travel demand away from small/medium airports to large hubs up to hundreds of miles away (Sixel Consulting Group Inc., 2014; Michigan Department of Transportation, 2016; Delta Airports Consultants Inc., 2020).

External shocks have further fueled air service disparities. The 2001 9/11 attacks and 2008 financial crisis led to partial to complete service losses (mainly due to the exit of airlines) at 30 and 23 small airports, respectively (General Accounting Office, 2002; US Department of Transportation, 2002; Hotle and Mumbower, 2021). Similar effects are also anticipated in the COVID-19 pandemic recovery, particularly at the end of the Coronavirus Aid, Relief, and

2.2 Airport Choice in the Context of Interregional Passenger “Leakage”

Air passengers’ choices for a given airport will be driven by the services offered at that airport, compared to the other(s) that are accessible to them, in addition to other socioeconomic traits and travel situations. As service disparities between airports grow, airports with better air services will attract more passengers from farther afield, sometimes up to hundreds of kilometers. It should be noted that studies of long-distance airport substitution differ from those of airport choice in urban multi-airport regions such as the San Francisco Bay Area and New York, in terms of ground access modes considered and data availability. Long-distance airport choice has been studied since the 1980s, including in the US states of California and Texas (Kanafani and Abbas, 1987; Kaemmerle, 1991), Northern New Brunswick, Canada (Innes and Doucet, 1990), Sheffield, UK (Thompson and Caves, 1993), and several Japanese cities (Furuichi and Koppelman, 1994). These studies show that air passengers access distant hub airports hundreds of kilometers away, to use jet services (over turboprop) and more frequent flights.

The phenomenon of long-distance passenger loss to larger airports was more commonly referred to as air passenger “leakage” starting in the early 2000s, with more studies in the US emerging since that time. Identified as major factors driving “leakage” include airfare (Suzuki, Crum, and Audino, 2003; Suzuki and Audino, 2003; Suzuki, Crum, and Audino, 2004; Phillips et al., 2005; Yirgu, Kim, and Ryerson, 2021), nonstop services, previous airport experience and frequent flier membership (Suzuki, Crum, and Audino, 2003), flight frequency and travel experience (Zhang and Xie, 2005), personal income and parking fee (Blackstone, Buck, and Hakim, 2006), travel duration and type of airline (Leon, 2011), and travel group and travel purpose (Gao, 2020). Furthermore, long-distance airport choice has also been studied extensively in both the European and Asian contexts within the last two decades. de Luca (2012) and Lian and Rønnevik (2011) show that air passengers in Southern Italy and Norway travel in excess of 2.5 and 4 hours, respectively, to access cheaper, nonstop flights. However, the European context differs from that of North America, given higher population densities and smaller distances between airports, and the frequent availability of extensive rail services. Air and rail competition (Jiang and Zhang, 2016), or the situation in which passengers access larger airports farther away using intercity rail
(instead of by private vehicle as per the typical North American context) (Albalate, Bel, and Fageda, 2015) has been studied extensively (Adler, Pels, and Nash, 2010; Dobruszkes, 2011).

2.3 Airport Catchment and Passenger Data Collection Approaches

The concept of air passenger “leakage” is predicated on the notion that airports have catchment areas – the land area from which an airport attracts passengers (Gao, 2020). Airports have air passengers that “belong” to them, demarcated by the geographic area from which they originate; in the case of “leakage,” these passengers are going further afield for air services. Various approaches have been presented to define airport catchment, including drawing circles of various radii around airports (Fuellhart, 2007; Fröhlich and Niemeier, 2011), and identifying areas within predetermined airport access times (Suzuki, Crum, and Audino, 2003; Marcucci and Gatta, 2011; Zhou et al., 2018; Milwaukee Mitchell Airport, 2020) and access distances (Suau-Sanchez, Burghouwt, and Pallares-Barbera 2014; Huber et al. 2021). In Europe, where nearly two-thirds of people live within two hours of at least two airports (Blachut, 2017), the various access time and distance thresholds used to define catchment boundaries result in considerable inconsistencies. State (Gao, 2020) as well as metropolitan area (Loo, Ho, and Wong, 2005; Teixeira and Derudder, 2021) boundaries are also used to demarcate airport catchments in previous studies.

While the aforementioned studies propose predefined and static catchment boundaries, Lieshout (2012) presents a different approach in a study of the region around Amsterdam Schiphol Airport. The author considers individual destinations and quality of air services at both the subject and surrounding airports in a basic multinomial logit model formulation (with parameters drawn from other sources rather than estimated based on a choice dataset), towards identifying areas where the subject airport has at least 1% passenger market share. The collection of these areas constitutes the catchment for each destination. Our study builds on this work by estimating a mixed-logit model on air tickets sold over a large portion of the US Midwest, where people have access to multiple airports offering a range of services.

Air passenger survey data has typically been collected towards studying airport “leakage.” Collection methods have included intercept surveys conducted at selected municipality centers (de Luca, 2012) and airports (Zhang and Xie, 2005), mail or telephone surveys (Blackstone, Buck, and Hakim, 2006) and combinations of the above (Suzuki, Crum, and Audino, 2003; Suzuki, 2007).
Other methods include purchased air ticket databases sampled from travel agents (Innes and Doucet, 1990; Leon, 2011; Yirgu, Kim, and Ryerson, 2021), and open access survey results (Lian and Rønnevik, 2011). Data differs significantly in geographic scope and passenger detail depending on the source. For instance, sampled air tickets often do not provide personal details although they are much larger in size and scope. Passenger surveys will include details such as travel purpose, travel experience, travel group, age, gender, income, airport access mode, frequent flier membership, and others. Despite their value, air passenger surveys over large regions are difficult if not impossible, as they would require substantial resources and public agency support is unlikely due to the complexities of coordinating across regional boundaries. Metropolitan Planning Organizations (MPOs) in the US, such as that of the San Francisco Bay Area, collect travel survey data that has supported studies of airport choice within multi-airport regions – such as Başar and Bhat (2004) and Hess and Polak (2005). But there has yet to be coordination amongst MPOs and neighboring regions towards collecting inter-regional travel survey data to study long-distance airport choice. Survey data to support passenger “leakage” have typically been collected and/or sponsored by airport authorities that target very specific areas from which passengers are suspected to “leak” to other airports. Consequently, geographic coverage and thus, more generalized insights into the drivers of “leakage,” are limited. To address this gap in the literature, our study uses a large sample of air tickets purchased by air passengers over a large portion of the US Midwest in an empirical airport choice model, to understand the main air service-based drivers of airport “leakage.”

3. Study Area

We study the departure airport choices of air passengers, based on the air tickets they purchased, originating from the US Midwest from January 2013 through December 2018. Major economic drivers in the region include manufacturing, education, health, tourism, and shipping and logistics (Siddiqui, 2022). Demand for air services in the region, during the above time period, was also stimulated by the entry of increasing numbers of Fortune 500 and 1000 companies (Airline Network News and Analysis, 2020). As shown in Fig. 1, the area we focus on includes parts of Illinois, Wisconsin, Iowa, Indiana and Michigan.
The study area centers on Chicago O’Hare International Airport (ORD), the third busiest airport in the US. In 2018, ORD had nonstop flights to over 210 domestic and international destinations, with a total of 39.87 million passenger enplanements (deplanement not considered here) (FAA, 2019). That same year, extensive infrastructure and capacity expansion programs including a new terminal, new gates and more concourses were started at an estimated budget of 8.5 billion USD (Ruthhart, 2018), with those programs still underway. ORD attracts air passengers from throughout Illinois and the neighboring states of Wisconsin, Indiana and even Michigan (Milwaukee Mitchell Airport, 2015; Naczek, 2019; Gao, 2020; Yirgu and Kim, 2021).

To the east of the study area is Detroit Wayne County Airport (DTW), which also draws passengers from adjacent states (Michigan Department of Transportation 2016; Yirgu, Kim, and Ryerson, 2021). Minneapolis St. Paul International (MSP), to the west, also attracts passengers from significant distances (Leon, 2011). Milwaukee Mitchell, Indianapolis and St. Louis Lambert International Airports are medium hub airports that offered nonstop services to various domestic destinations (approximately 30-60) during the study period of 2013-2018. The remaining small hub and non-hub primary airports in the study area offered far fewer daily flights (two to five per
destination, on average) to a number of large hubs both within and beyond the study region. LCCs provided point-to-point services at some of these small and non-hub airports.

4. Data Sources and Descriptions

4.1 Air Tickets

Market Locator is a dataset of air ticket purchases made from online and traditional brick-and-mortar travel agencies. We acquired a sample in May 2019 from the Airlines Reporting Corporation (ARC), for the study time period and geographic area introduced previously. Each complete record in Market Locator consists of origin airport, destination airport, ticketing airline, route flown, number of passengers, month and year of flight, and billing ZIP code of the credit card used to purchase the ticket. It does not include passenger-specific attributes such as travel purpose, travel experience, travel group, age, gender, income, airport access mode, frequent flier membership etc. The dataset predominantly represents leisure travelers (Drukker, 2022).

4.2 Air Service Characteristics

Airfare, flight frequency, availability of nonstop services, and airport accessibility are among the most documented transportation service attributes affecting departure airport choice, as discussed in the literature review, and we extract these from various sources maintained by the US Department of Transportation, Bureau of Transportation Statistics. The Market Locator dataset does not report airfare charged per itinerary record, but airfares for similar itineraries and timeframes are available in the Airline Origin and Destination Survey (DB1B) (BTS, n.d.a). For flight frequency, we use the T-100 Domestic Segment (All Carriers) dataset of Air Carriers Statistics (Form 41 Traffic) (BTS, n.d.b).

4.3 Data Filtering and Processing

In several instances, more than one passenger (sometimes up to hundreds) is reported on a single ticket record in the Market Locator dataset. In datasets such as DB1B, it is common to report passengers that do not necessarily belong to the same travel group together for the purposes of compressing datasets (Martin, Martin, and Lawford, 2010). In the absence of other information that could help identify air passengers that traveled together, we assume each traveler is an independent decision maker, thereby excluding a travel group attribute.
Records of passengers in the Market Locator dataset are supplemented with mean airfare from DB1B and total flight frequencies computed from T-100, by quarter-year. Unlike nonstop flights, the frequencies of flights between an origin airport and the first stop airport are considered for connecting itineraries, as these primarily determine the schedule and subsequent connection time of passengers (Marcucci and Gatta, 2011). While extracting these attributes from DB1B and T-100, itineraries on Southwest Airlines are excluded as the Market Locator underrepresents tickets from this carrier (it has a 95% market share at Midway International, Chicago’s other large hub, and thus Midway was excluded from this study) and the smaller LCCs of Allegiant Air and Spirit (Sixel Consulting Group Inc., 2014).

Finally, the ZIP codes associated to the credit cards used to purchase air tickets are assumed to represent passengers’ home addresses, a reasonable assumption given that the dataset predominantly consists of leisure travelers (Drukker, 2022). We compute the distances between the geographic coordinates of the centroids of the reported ZIP codes (Opendatasoft, n.d.) and airports (IATA, n.d.), along a network of primary and secondary state roads (US Census Bureau, n.d.), to represent passenger access distances. This is performed in ArcMap using the Network Analyst tool. We note our use of access distance instead of access time as a measure of airport accessibility, because the latter can vary significantly by day of week and time of day.

5. Analysis Approach

The analysis approach taken in this study is as follows. First, a mixed logit model (MMNL) is used to estimate how air service characteristics drive air passengers’ airport choices, specifically on routes originating from the 21 departure study airports (Fig. 1) and ending at 61 unique final destinations (Appendix A). MMNL captures heterogeneity across decision-makers (air passengers, in our case), through estimated model coefficients that are random variables rather than deterministic constants that result from multinomial logit (MNL) and nested logit (NL) models. Model specification and estimation are discussed in Section 5.1. The 61 final destinations are recorded to have at least one passenger traveling to each from each of the study airports. This approach assumes that all air passengers within the study area can choose among all 21 airports. Such an assumption has some limitations given that the airports considered by a passenger are limited and depend, among other things, on sociodemographic, informational, psychological and subjective criteria specific to that person (Ben-Akiva and Boccara, 1995; Başar and Bhat, 2004).
As such, an individual might decide a particular airport to be too far (or too unattractive because it is not served by an airline they frequently use) to even consider, while another individual might not (Başar and Bhat, 2004). Given the lack of data on the abovementioned factors, we use airport access distance combined with the air service characteristics offered at each airport to explain airport choices. The routes are classified into short, medium, and long total flight length categories. The resulting airport choice characteristics are compared across these categories to understand how air service attributes influence airport choices. The categorization is described in Section 5.2 and the results are presented in Section 6.1.

Next, for each of the 21 departure airports, hypothetical destinations are created by using averages of the air service attributes on all 61 destinations. Model parameters are estimated again, and the results are plotted for selected airports (discussed in Section 6.2). Finally, the effects of disparities in air service characteristics, on airport market areas and shares, are explored through a sensitivity analysis (discussed in Section 6.3).

5.1 Model Estimation

We presume airport choices are driven by mean airfare, airport access distance, flight frequency and type of service (nonstop vs. connecting). Other service-based explanatory variables such as fare per mile, seats offered, number of flight legs and route distance were also initially explored. Different combinations of these variables were tested for multicollinearity using correlation matrices and variance inflation factor (VIF). For VIF, an upper limit of ten is applied to limit non-pairwise correlation (O’Brien, 2007), such that any combination of variables whose VIF value exceeds ten is judged to exhibit collinearity. Finally, the combination of variables shown in Equation 1 is retained for the utility function.

\[
U_{i,j,k}^q = \gamma_{i,j,k} + \epsilon_{i,j,k} = \alpha_{0,k}I(i) + \alpha_{1,k}I_{i,j}^q + \alpha_{2,k}D_{i,k} + \alpha_{3,k}F_{i,j}^q + \alpha_{4,k}NS_{i,j}^q + \epsilon_{i,j,k}
\]  

(1)

Where

\( q = \) quarter-year, \( q = 1 \ldots Q \), where \( Q \) is 24 (4 quarters per year for 6 years, 2013-2018);

\( i = \) departure airport, from \( i = 1 \ldots 21 \) airports;

\( j = \) final destination airport, \( j = 1 \ldots 61 \) airports;
\( k \) = decision maker (air passenger), \( k = 1 \ldots K \);

\( U_{i,j,k}^q \) = is the utility of choosing the route from departure airport \( i \) to final destination airport \( j \) for passenger \( k \), in quarter-year \( q \);

\( V_{i,j,k}^q \) = is the observed utility on route \( i \) to \( j \) for \( k \), in \( q \);

\( I(i) = \) airport dummy, \( I = 1 \) if departure airport chosen is \( i \), and 0 otherwise;

\( f_{i,j}^q \) = mean airfare (USD) on route \( i \) to \( j \) in \( q \);

\( D_{i,k} \) = access distance (miles) from residential ZIP code of \( k \) to \( i \);

\( F_{i,j}^q \) = flight frequency on route \( i \) to \( j \) in \( q \);

\( NS_{i,j}^q \) = nonstop service dummy on route \( i \) to \( j \) in \( q \), \( NS_{i,j}^q = 1 \) if flight is nonstop, 0 if connecting;

\( \alpha_{0,k} \) = alternative specific constant (ASC) for \( k \);

\( \alpha_{1,k}, \alpha_{2,k}, \alpha_{3,k}, \alpha_{4,k} = \) coefficients on air service attributes for \( k \), and

\( \varepsilon_{i,j,k}^q \) = error term on route \( i \) to \( j \) for \( k \), in \( q \).

Again, the Market Locator dataset does not include information such as passenger-specific attributes. Therefore, the utility of alternatives cannot be (and regardless, is never) fully captured, resulting in randomly distributed error terms or unobserved portions of utilities (Bernasco, Wim; Block, 2009). However, by introducing alternative specific constants (ASCs), the mean of the error terms can be added to the observed utility function, such that the remaining normally distributed error term has a mean of zero (Hess and Polak, 2005).

The MMNL treats parameters \( \alpha_0 \ldots \alpha_4 \) as random variables rather than constants, and thus model estimates are of the moments of these parameters. If they are assumed to be normally distributed, their means and standard deviations are estimated. As shown in Equation 2, the probability that decision maker \( k \) chooses departure airport \( i \) to fly to destination \( j \) during quarter-year \( q \), i.e. \( P_{i,j,k}^q \), is an integral of the conditional logit probability over some density distribution \( f(\alpha|\Omega) \), where \( \Omega \) represents the parameters of this distribution.

\[
P_{i,j,k}^q = \int P_{i,j,k}^q(\alpha) \cdot f(\alpha|\Omega) \, d\alpha 
\] (2)

and
\[ L_{i,j,k}^q(\alpha) = \frac{e^{V_{i,j,k}^q}}{\sum_{l=1}^L e^{V_{l,j,k}^q}} \]  

Where

\[ L_{i,j,k}^q(\alpha) = \text{probability that } k \text{ chooses } i \text{ to fly to } j \text{ in } q, \text{ conditional on } \alpha, \text{ and other variables as previously defined.} \]

Equation 2 has no closed form solution and is solved by simulation. While various distribution types are possible, the normal distribution is chosen as it does not force only positive parameter estimates (Ibeas et al., 2014) like, for example, the log-normal distribution. When only positive or negative estimates are forced, the researcher must decide the signs of the estimated parameters a-priori (Hess and Polak, 2005).

Equation 3 is approximated through simulation for any given value of \( \Omega \) using the following steps:
i) draw a value of \( \alpha \) from \( f(\alpha|\Omega) \) to be labeled as \( \alpha^g \) where \( g = 1 \) refers to the first draw; ii) calculate \( L_{i,j,k}^q(\alpha^g) \) using Equation 3, and iii) repeat steps i and ii a total of \( G \) times (number of simulations), and average results. This average is the simulated probability shown in Equation 4.

\[ \hat{P}_{i,j,k}^q = \frac{1}{G} \sum_{g=1}^G L_{i,j,k}^q(\alpha^g) \]  

For the entire set of decision makers \( K \), the value of \( \Omega \) that maximizes the simulated log-likelihood (SLL) function is found (McFadden and Train, 2000) using Equation 5.

\[ SLL = \sum_{k=1}^K \sum_{l=1}^L d_{i,j,k}^q \hat{P}_{i,j,k}^q \]  

Where

\[ d_{i,j,k}^q = \text{dummy variable on } \hat{P}_{i,j,k}^q, \text{ 1 if } k \text{ chooses } i \text{ to travel to } j \text{ in } q, \text{ and 0 otherwise.} \]

The coefficient of variation (CV), the ratio of standard deviation to mean (Li, 2015), is a dispersion indicator used to measure decision-maker taste heterogeneity with respect to the attributes.
5.2 Segmentation by Flight Length

Air passengers can be categorized into more homogeneous groups, and model estimates across these different groups can be compared. In previous studies, passengers have been categorized by trip purpose (business vs. leisure) and geography familiarity (resident vs. non-resident) (Hess and Polak, 2005; Ishii, Jun, and Van Dender, 2009), amongst others. In this work, we follow previous work in categorizing passengers based on total flight distance traveled (Jorge-Calderón, 1997; Mohammadian et al., 2019; Lee, Li, and Song, 2019), to understand whether the influences of air service attributes on different flight distances vary and by how much. Based on recommendations for domestic US operations (Hansen, 1995) and Eurocontrol’s standard classification (European Organisation for the Safety of Air Navigation, 2011), three categories are chosen: <932 miles (short route), 932-1864 miles (medium route), and >1864 miles (long route).

5.3 Passenger Sample Subset

As expected, initial attempts to implement the specified MMNL on all 2.37 million passengers in the three flight range categories did not lead to convergence of the SLL function. Although simpler model structures that converge more easily such as MNL and NL were tested with good results, they do not offer the random variable coefficients sought after. It was deemed important to capture that decision makers are unlikely to place the same value on an attribute (Hess and Polak, 2005). Second, certain airports are likely to be viewed as more substitutable by passengers, particularly when similar airlines serve these kinds of airports, which is not accounted for with a MNL model.

To achieve convergence with a smaller dataset, a subset of the total passenger dataset is drawn using Halton sequences which, when compared to simple random sampling, improves the spatial coverage of geographical data (Robertson et al., 2019). We first fix the minimum subset data size of the base category, chosen to be all passengers in the long route category. For population sizes over 100,000, a sample size of 2,000 is considered sufficient (Koppelman and Chu, 1983; de Bekker-Grob et al., 2015). Based on this, a sample size of 2,000 passengers is taken for the long route category. Once this category’s size is fixed, the subset sizes of the remaining two categories are fixed by preserving the proportion ratio observed in the total dataset, as typically done in clustered sampling (Meng, 2013; IBM, 2021). For instance, the ratio of passengers on short routes to those on long routes is 2.12, while that of short to medium is 0.39 in the processed ticket dataset.
The ratio of passengers on medium routes to those on long routes is 5.5. Accordingly, subset sizes of 4,240, 11,000 and 2,000 passengers are chosen for the short, medium and long route categories, respectively. Given that only subsets of the total dataset are used to estimate parameters, the estimates will vary depending on the data subset used. Thus, for each route category, five different subsets are drawn to check if model estimates exhibit statistically significant differences.

6. Results and Discussion

6.1 Route Distance-based Model Estimates

The model estimates resulting from the five data subsets were compared using a t-test. Differences in model estimates for airfare, flight frequency, access distance and nonstop service between two data subsets were found to be statistically significant. Also, differences in ASC estimates on three of the 21 airports, i.e., MSP, DSM and STL, were found to be statistically significant between all data subsets. However, the comparisons among estimates on the different route categories discussed hereafter still hold for all data subsets. With this, one data subset per route category is randomly chosen, and the estimated parameter results are discussed. The means and standard deviations of the ASCs have been excluded in the interest of length, but are included in Appendix B. Parameters are estimated using the mlogit library of R and presented in Table 1. All estimates are significant at the 99% confidence level, with standard errors indicated in parenthesis next to model estimates.
Table 1 MMNL Model Estimates

<table>
<thead>
<tr>
<th>Utility variable</th>
<th>Moments</th>
<th>Short routes</th>
<th>Medium routes</th>
<th>Long routes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Estimate (std. error)</td>
<td>Mean</td>
<td>Estimate (std. error)</td>
</tr>
<tr>
<td>Access distance</td>
<td>-0.102</td>
<td>(0.009)</td>
<td>-0.107</td>
<td>(0.006)</td>
</tr>
<tr>
<td></td>
<td>Std. dev</td>
<td>0.046 (0.005)</td>
<td>0.053</td>
<td>(0.003)</td>
</tr>
<tr>
<td></td>
<td>CV, %</td>
<td>45.1</td>
<td>49.5</td>
<td>38.8</td>
</tr>
<tr>
<td>Flight frequency</td>
<td>Mean</td>
<td>0.001 (0.000)</td>
<td>0.001</td>
<td>(0.000)</td>
</tr>
<tr>
<td></td>
<td>Std. dev</td>
<td>0.000 (0.000)</td>
<td>0.001</td>
<td>(0.000)</td>
</tr>
<tr>
<td></td>
<td>CV, %</td>
<td>0.0</td>
<td>100.0</td>
<td>166.7</td>
</tr>
<tr>
<td>Mean airfare</td>
<td>Mean</td>
<td>-0.039 (0.004)</td>
<td>-0.039</td>
<td>(0.003)</td>
</tr>
<tr>
<td></td>
<td>Std. dev</td>
<td>0.029 (0.004)</td>
<td>0.029</td>
<td>(0.004)</td>
</tr>
<tr>
<td></td>
<td>CV, %</td>
<td>74.4</td>
<td>74.4</td>
<td>176.2</td>
</tr>
<tr>
<td>Nonstop service</td>
<td>Mean</td>
<td>2.922 (0.370)</td>
<td>2.625</td>
<td>(0.180)</td>
</tr>
<tr>
<td></td>
<td>Std. dev</td>
<td>1.860 (0.493)</td>
<td>2.126</td>
<td>(0.239)</td>
</tr>
<tr>
<td></td>
<td>CV, %</td>
<td>63.7</td>
<td>81.0</td>
<td>168.8</td>
</tr>
</tbody>
</table>

Passengers sampled in data subset: 4,240, 11,000, 2,000
Observations: 89,040, 231,000, 42,000
SLL: -3,592.317, -9,198.864, -1,559.154
McFadden $R^2$: 0.733, 0.723, 0.744
Adjusted McFadden $R^2$: 0.730, 0.722, 0.736

All parameter estimates are significant at the 99% confidence level.

Coefficients of variation (CVs) exceeding 100% are often associated with dispersion (Bedeian and Mossholder, 2000), and are indicative of greater taste heterogeneity. The CV of access distance ranges between 38.8% and 49.5% while that of frequency ranges between 0% and 166.7% for the different route categories. Additionally, the CV of airfare for both short and medium routes is 74.4%, but increases to 176.2% for long routes. Nonstop service has a CV ranging between 63.7% for short routes up to 168.8% for long routes.

Results show that the influences of airfare, flight frequency and access distance on airport choice are very similar for short and medium routes. Airfare mean coefficients are -0.039 for both, while access distance mean coefficients are -0.102 and -0.107 for short and medium routes, respectively (about a 5% difference). The influence of nonstop service on airport choices differs at 2.92 for short vs 2.63 for medium routes (~10% difference).

The impacts of airfare and nonstop service on air passengers’ utilities on long routes differ compared to those on short and medium routes (~0.042 for long versus ~0.039 for short/medium
on airfare). For comparable airfares, the utility portion of airfare on long routes is expected to be 7% lower than those on short/medium. However, for comparable flight frequencies, passengers on long routes place three times as much weight (0.003 vs. 0.001) on this frequency attribute compared to those on short and medium routes. These results are not straightforward to interpret. For nonstop routes, the frequency considered is the number of flights serving those nonstop routes. For connecting routes, which are often associated with longer flights, only the first leg flight frequencies are considered. In instances where these connecting routes are served through multiple intermediate airports, the means of flight frequencies leading to these intermediate airports are taken. Marcucci and Gatta (2011) showed that connecting air passengers attempt to minimize their wait times during connections, which is in large part determined by the first flight schedule, as more options usually exist at airports where connections are made. Additionally, even when nonstop services are available, long routes are usually offered at lower frequencies than short routes. Thus, passengers taking long routes are likely to place greater importance on flight schedule, with lower flight frequencies further restricting passengers’ schedule preferences. Passengers on long routes, compared to those on short ones, place twice as much weight on access distance (-0.201 vs -0.102 for short-range), which seems counterintuitive as passengers flying to farther destinations would be expected to be more amenable to more distant departure airports, because of an inclination to value time and cost in relative terms. Instead, these results suggest a possible inclination towards minimizing overall travel time, which includes the ground-based airport access trip plus the longer flight. Longer trips are associated with increased value of travel time (Athira et al., 2016; US Department of Transportation, 2016), given that passengers’ abilities to engage in other productive activities decrease as they spend more time traveling. Thus, they are willing to pay more to save a unit of time.

The relatively high CVs associated with airfare (176.2), frequency (166.7) and nonstop service (168.8%) on long routes imply that passengers’ responses to these attributes are more diverse. In essence, we may be able to better pinpoint how changes in these service attributes would affect passengers on short and medium routes. To further investigate, the distributions of the model coefficients are presented in Figs. 2–3.
Fig. 2 Estimated parameter distributions for flight frequency (left) and nonstop service (right)
From Fig. 2, 11% and 12% of air passengers on short and medium routes, respectively, have negative coefficients for flight frequency. This indicates that higher flight frequency is expected to increase the probability of an airport being chosen by the remaining 89% and 88% of passengers. Similarly, the availability of nonstop services has a positive influence on an airport’s probability of being chosen among 94% and 89% of passengers on short and medium routes, respectively. These findings confirm that higher flight frequency and nonstop service availability lead to a positive preference response across the majority of passengers on short and medium route. However, for passengers on long-routes, both flight frequency and nonstop services positively influence an airport’s choice probability among a smaller 76% and 73% of passengers, respectively. The estimated values on airfare and access distance are shown in Fig. 3.
Fig. 3 Estimated parameter distributions for airport access distance (left) and airfare (right)
The left column graphs of Fig. 3 show that 97%-99% of passengers across the three route categories are negatively influenced by greater airport access distance, as expected, confirming a preference for closer airports. Higher airfares have a negative influence on 90% and 91% of passengers on short and medium routes, respectively. This drops to 71% for passengers on long routes, meaning with all else equal, 29% of these passengers view paying more for airfare positively towards choosing an airport. Although counterintuitive, this result might suggest that there are other attributes – such as preferences for certain airlines that service these routes, or tickets that include the costs of add-ons – that have passengers choosing long-route itineraries with higher airfares.

6.2 Airport Choice Probabilities by Geography

Airport choice probabilities are calculated using the means of the estimated model parameters applied to hypothetical destinations. These hypothetical destinations feature the mean air service attributes for traveling to the 61 destinations identified previously. This is done as an alternative to showing probability plots for each of the 61 destinations individually, and serves the purpose of demonstrating how air service disparities impact passengers’ airport choices across the study region. The mean air service attributes are shown for all studied departure airports, by airport size classification, in Fig. 4.
Fig. 4 Mean air service attributes representing hypothetical destinations by departure airport size category

From Fig. 4, mean airfare and nonstop service are highest and lowest, respectively, at the non-hub primary airports, as expected. Large hubs have mean nonstop service indicators at or near one (1 at ORD, 0.8 at DTW, 0.9 at MSP), as most of the 61 destinations were directly connected to these airports. Medium hubs have nonstop service indicators from 0.35-0.45, small hubs from 0.16-0.24, and all non-hubs less than 0.15. Flight frequency is computed to consider the schedule of first leg flights for destinations not served nonstop, from all non-hub primary and small hubs to the 61 destinations. As discussed earlier, the characteristics of first leg flights are critical in passengers’ airport and itinerary choices towards reducing wait times but increasing the likelihood of catching connecting flights (Marcucci and Gatta, 2011).

Because attributes representing mean air service characteristics across all destinations are considered here, the model in Section 5.1 is first estimated for all 61 destinations without the route distance categorization of Section 5.2. The minimum sample size of 2,000, Halton sequencing mechanism and five different data subset draws employed in 5.3 are used here as well.\(^1\) Model

---

\(^1\)Higher sample sizes that achieve convergence were tested, and a final size of 17,000 is used. However, for the flight distance-based sampling, it was not possible to increase sample sizes while still preserving cluster ratios and achieving convergence.
estimates, with standard errors in brackets, are shown in Table 2 based on one data subset. Then, these estimates are applied on the air service attributes of the hypothetical destinations, to find each departure airport’s probability of being chosen, by ZIP code as discussed hereafter.

Table 1 MMNL Model Estimates without Route-based Categories

<table>
<thead>
<tr>
<th>Utility variable</th>
<th>Mean (std. error)</th>
<th>Std. dev (std. error)</th>
<th>CV, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access distance</td>
<td>-0.095 (0.004)</td>
<td>0.037 (0.002)</td>
<td>38.9</td>
</tr>
<tr>
<td>Flight frequency</td>
<td>0.001 (0.000)</td>
<td>0.001 (0.000)</td>
<td>100.0</td>
</tr>
<tr>
<td>Mean airfare</td>
<td>-0.033 (0.002)</td>
<td>0.026 (0.003)</td>
<td>78.8</td>
</tr>
<tr>
<td>Nonstop service</td>
<td>1.765 (0.113)</td>
<td>1.839 (0.176)</td>
<td>104.2</td>
</tr>
<tr>
<td>$I(\text{ATW} = 1)$ (base)</td>
<td>0 (-)</td>
<td>0 (-)</td>
<td></td>
</tr>
<tr>
<td>$I(\text{AZO} = 1)$</td>
<td>-1.155 (0.435)</td>
<td>0.189$^\text{ns}$ (0.828)</td>
<td>16.4</td>
</tr>
<tr>
<td>$I(\text{BMI} = 1)$</td>
<td>2.740 (0.302)</td>
<td>2.437 (0.388)</td>
<td>88.9</td>
</tr>
<tr>
<td>$I(\text{CMI} = 1)$</td>
<td>-0.547$^\text{**}$ (0.320)</td>
<td>0.251$^\text{ns}$ (0.505)</td>
<td>45.9</td>
</tr>
<tr>
<td>$I(\text{CWA} = 1)$</td>
<td>-1.829 (0.332)</td>
<td>0.510$^\text{ns}$ (0.682)</td>
<td>27.9</td>
</tr>
<tr>
<td>$I(\text{DBQ} = 1)$</td>
<td>-4.305 (0.501)</td>
<td>2.725 (0.632)</td>
<td>63.3</td>
</tr>
<tr>
<td>$I(\text{DSM} = 1)$</td>
<td>5.357 (0.439)</td>
<td>5.063 (0.487)</td>
<td>94.5</td>
</tr>
<tr>
<td>$I(\text{DTW} = 1)$</td>
<td>6.799 (0.393)</td>
<td>1.214 (0.243)</td>
<td>17.9</td>
</tr>
<tr>
<td>$I(\text{FWA} = 1)$</td>
<td>2.243 (0.355)</td>
<td>0.215 (0.534)</td>
<td>9.6</td>
</tr>
<tr>
<td>$I(\text{GRB} = 1)$</td>
<td>0.282$^\text{*}$ (0.142)</td>
<td>1.319 (0.259)</td>
<td>46.7</td>
</tr>
<tr>
<td>$I(\text{GRR} = 1)$</td>
<td>3.079 (0.297)</td>
<td>2.725 (0.272)</td>
<td>88.5</td>
</tr>
<tr>
<td>$I(\text{IND} = 1)$</td>
<td>7.776 (0.401)</td>
<td>1.905 (0.271)</td>
<td>24.5</td>
</tr>
<tr>
<td>$I(\text{MKE} = 1)$</td>
<td>5.808 (0.316)</td>
<td>3.702 (0.317)</td>
<td>63.7</td>
</tr>
<tr>
<td>$I(\text{MLI} = 1)$</td>
<td>1.974 (0.264)</td>
<td>3.888 (0.521)</td>
<td>197.0</td>
</tr>
<tr>
<td>$I(\text{MSN} = 1)$</td>
<td>2.589 (0.241)</td>
<td>2.320 (0.327)</td>
<td>89.6</td>
</tr>
<tr>
<td>$I(\text{MSP} = 1)$</td>
<td>9.229 (0.493)</td>
<td>4.147 (0.535)</td>
<td>44.9</td>
</tr>
<tr>
<td>$I(\text{ORD} = 1)$</td>
<td>6.832 (0.419)</td>
<td>2.410 (0.316)</td>
<td>35.3</td>
</tr>
<tr>
<td>$I(\text{PIA} = 1)$</td>
<td>3.883 (0.425)</td>
<td>1.543 (0.743)</td>
<td>39.7</td>
</tr>
<tr>
<td>$I(\text{SBN} = 1)$</td>
<td>1.838 (0.352)</td>
<td>0.504$^\text{ns}$ (0.481)</td>
<td>27.4</td>
</tr>
<tr>
<td>$I(\text{SPI} = 1)$</td>
<td>-0.918$^\text{ns}$ (0.926)</td>
<td>0.722$^\text{ns}$ (0.959)</td>
<td>78.6</td>
</tr>
<tr>
<td>$I(\text{STL} = 1)$</td>
<td>10.658 (0.576)</td>
<td>5.751 (0.685)</td>
<td>54.0</td>
</tr>
</tbody>
</table>

Passengers sampled in data subset 17,000
Observations 357,000
SLL -14,906.384
McFadden R² 0.718
Adjusted McFadden R² 0.717

*significant at 95% level; **significant at 90% level; n/s not significant at 90% level. All others significant at 99%.

In the interest of length, the discussion is focused on choice probability plots for one airport each from the small, medium and large categories: Dane County Regional (MSN), Milwaukee Mitchell...
International (MKE), and Chicago O’Hare International (ORD), respectively. ORD’s position as one of the largest US hub airports presents an opportunity to show the geographic extent from which an airport can attract passengers. MKE and MSN are also chosen due to their proximity and documented history of passengers within their assumed catchments “leaking” to ORD (Milwaukee Mitchell Airport, 2015; Yirgu, Kim, and Ryerson, 2021).

Figs. 5–6 show the airport choice probabilities for MSN, MKE and ORD after applying the mean values of model parameter estimates (Table 2) on the mean air service attributes of Fig. 4. These airport choice probabilities can be interpreted as, or represent, the airport market shares (Wei and Hansen, 2005; Ryerson and Kim, 2018). State borders as well as MPO boundaries are shown. All ZIP codes with airport market shares of 0.05 and larger are included. Note that results are restricted to the study ZIP codes of Fig. 1. The coverage area of all plots are identical to allow for visual comparisons between results.

---

Fig. 5 Market areas and shares of MSN (left) and MKE (right)
Several observations are made from Figs. 5–6. First, as airport size increases, its market area also increases. As expected, MSN has the smallest market area followed by MKE, while ORD has by far the largest and strongest. Second, the area over which ORD has complete market dominance (0.98–1.00) is larger than that of MKE, while MSN does not have market shares exceeding 0.95 in any ZIP code. Third, small and medium airports’ market areas and shares are smaller in the direction towards a neighboring airport, indicating that these airports are more susceptible to airport competition. For instance, the market area of MSN mainly extends to its west (with no dominant airports) as MKE is located to its east and ORD to its southeast. Similarly, MKE’s market area extends primarily north, as ORD is located only 72 mi to the south and as such, passengers located south of MKE are far more likely to access ORD. Although results are not shown here, similar observations can also be made for the other study airports.

Administrative borders are, as expected, of no importance when it comes to airport passenger markets. The area over which MSN’s market shares exceed 0.90 crosses seven MPO boundaries, while ORD’s crosses 21. This does, however, raise attention to some critical surface transportation planning issues. In the US, where intercity transit services are the exception rather than the norm (Sperry et al., 2012; Augustin et al., 2014), air passengers throughout the region reach airports
mainly by private vehicle, potentially accounting for up to 2.75% of average daily traffic on congested portions of interstate highways (Ryerson and Kim, 2018). MPOs, which are responsible for regional transportation planning and are mainly structured to function independently (Rahman, Sciara, and Ryerson, 2021), focus on air transport demand only within their respective jurisdictional boundaries (Hess and Polak, 2005; 2006; Regional Airport Planning Committee, 2011; Steer Davies Gleave and Mark Kiefer Consulting, 2014). Aside from limited coordination with their neighboring counterparts, these MPOs also have little to no collaboration with the operators of several airports within their region. For example, Sciara (2019) reports that only 11% of MPOs provide voting seats for airports and an even lower 2.5% have dedicated aviation committees. Particularly in the case of long-distance travel behavior, this fractured planning paradigm cannot account for the fact that travelers originating from one jurisdiction frequently use surface and air transportation facilities and services in another.

The results of Figs. 5–6 also have implications for catchment-based airport demand estimates. The results clearly show that airport catchments cannot be demarcated based on circles of predetermined radii around airports (Fuellhart, 2007; Fröhlich and Niemeier, 2011; O’Connor and Fuellhart, 2016), areas within predetermined airport access times (Suzuki and Audino, 2003; Marcucci and Gatta, 2011; Zhou et al., 2018; Milwaukee Mitchell Airport, 2020) or distances (Suau-Sanchez, Burghouwt, and Pallares-Barbera, 2014; Huber et al., 2021), and administrative boundaries (Loo, Ho, and Wong, 2005; Gao, 2020; Teixeira and Derudder, 2021). Nonetheless, such definitions have extensively been used in airport planning and demand studies (Moore and Soliman, 1982; Rengaraju and Thamizh Arasan, 1992; Zhou et al., 2018; Mohammadian et al., 2019). Only Lieshout (2012) explicitly accounted for air service qualities in defining an airport’s catchment, while Adler et al. (2022) have more recently used a large dataset of anonymized GPS records generated by mobile application users to support airport catchment estimation. While there is general consensus that a one-size-fits-all catchment definition is insufficient (Gao, 2020; Huber et al., 2021), a lack of large-scale passenger data has largely stymied empirically-driven work in this area.

6.3 The Impacts of Air Service Disparities on Airport Markets

At airports, new services are launched or airlines exit certain markets due to various forces impacting passenger demand, including economic growth or decline, pilot shortage, market
profitability, disruptive events (such as the COVID-19 pandemic), and others (Fuellhart et al., 2016; Atallah and Hotle, 2019; Hotle and Mumbower, 2021). Such changes will not result in the same passenger responses at all airports. The following analysis provides some insights into how airports of different sizes and locations, in relation to other airports, may be impacted.

The effects of widening disparities in air services between airports, as represented by the mean air service attributes of Fig. 4, are investigated. These disparities are simulated by changing the mean value of one attribute at one airport, keeping all else equal across other attributes and airports. We first investigate the effects of increasing the mean airfare at MSN, MKE and ORD in 10% increments. The results for MSN are shown in Fig. 7, while those for MKE and ORD are shown in Figs. 8–9, respectively.

Fig. 7 Changes in airfare and resulting market changes, MSN (mean airfare = 287 USD)
As airfare increases at MSN, the area over which the airport has market share exceeding 0.05 contracts in all directions while market shares themselves also decrease. For instance, market shares are 0.90-0.95 at mean airfare in ZIP codes clustered around the airport, and these decrease to 0.70-0.90 and 0.50-0.70 at 20% and 30% higher mean airfares, respectively.

Fig. 8 Changes in airfare and resulting market changes, MKE (mean airfare = 264 USD)

From Fig. 8, contraction in market area and changes in market shares are observed for MKE much like MSN. However, Fig. 9 shows that such contractions are much less observable for ORD, reinforcing previous findings that large airports can typically absorb service degradations without losing substantial market (Ryerson, 2016; Atallah and Hotle, 2019). In general, MKE and ORD –
bigger and better serviced airports than MSN – do not have their market shares decrease by more than half over a substantial proportion of their market areas like MSN, owing to the larger ASCs estimated (5.8 for MKE and 6.8 for ORD, versus 2.6 for MSN). These ASCs capture the advantages larger, better serviced airports have over other those with worse/less air services, as well as characteristics difficult to quantify such as airport reputation (Hess and Polak, 2005). As a result, the impacts of utility reductions caused by deteriorations in air service attributes are less impactful on passenger utility for the subject airport. Ishii, Jun, and Van Dender (2009) also showed that the ASCs of larger airports that compete with smaller ones are higher, such that if air service attributes explicitly defined in the MMNL model and airport ground access distance are not accounted for, passengers would still prefer to depart from larger airports.

Fig. 9 Changes in airfare and resulting market changes, ORD (mean airfare = 280 USD)
The impacts of decreasing mean flight frequency and mean nonstop service on study airports’ market areas and shares are considerably smaller than those resulting from proportional percentage changes in airfare. Mean flight frequency and nonstop service at each subject airport were reduced in the same 10% increments up to 30%, while keeping all else at mean values. The greatest reductions in market shares are just under 15% and 25% for flight frequency and nonstop service, respectively. The same 30% increase in airfare led to market share reductions exceeding 75%.

Fig. 10 shows the reductions in market shares for MKE resulting from proportional increases in mean airfare (left column) versus decreases in mean flight frequency (right column). It is noted that only ZIP codes with choice probabilities of at least 0.05 after air service attribute changes are shown. Reduced mean flight frequencies have a much weaker effect on market shares compared to proportionally increased mean airfares. Moving farther away from MKE, the reduction in market share is considerably higher for increased airfare, reaching >75%, whereas that for decreased mean flight frequency peaks at just under 15%. While the above observations are based on comparable percentage changes in the different service attributes, it should be noted that air passengers internalize airfare and other air service attributes such as frequency differently.

While flight frequency has been shown to be critical in driving airport choices within multi-airport regions (Harvey, 1987; Windle and Dresner, 1995; Başar and Bhat, 2004; Hess and Polak, 2005), its importance has been reported to be limited or statistically insignificant in driving long-distance airport choices and passenger “leakage” (Yirgu, Kim, and Ryerson, 2021; Suzuki, Crum, and Audino, 2003). This may be due to the fact that leaking passengers are predominantly made up of leisure travelers, who are more likely to prioritize airfare savings over flight schedules.
Fig. 10 Market share reductions (%) at MKE caused by increases in airfare (left) and decreases in flight frequency (right)
7. Conclusion

Disparities in air services among airports across North America have deepened considerably over the last two decades and continue to do so, leading air passengers to make longer drives to access airports offering better air services. Using a large dataset of air tickets sales, we are able to abandon market catchment assumptions previously relied upon to study this phenomenon, towards assessing the departure airport choices of air passengers from a large portion of the US Midwest spanning multiple city, state and regional boundaries.

Using a mixed logit model, we observed greater response heterogeneity to well-established air service attributes such as airfare, flight frequency and nonstop services among passengers on long routes, compared with those on short and medium ones. Changes in air service attributes for short- and medium-length routes at an airport will have a larger influence on air passengers’ airport choices, compared with long routes. Among the air service attributes considered, airfare is found to impact airports preferences more than flight frequency and nonstop services. Comparable changes in air service attributes lead to greater changes in the market shares and thus, catchment areas, of small and medium airports, compared with large hubs. The results clearly show that an airport’s catchment area – specifically, small- and medium-size airports – is highly influenced by the proximity and characteristics of other neighboring airports. Furthermore, the area over which a large hub airport has very strong market shares extends strongly across administrative and geographic boundaries of multiple jurisdictions.

The results have a number of implications. First, small- and medium-size airports aiming to retain or increase market shares should focus on attracting airlines that provide services on short- and medium-length routes compared to long routes. Second, the market share/catchment plots confirm the strong draw of passengers to larger hub airports across an expansive geography, to areas well beyond which the airport authority, or MPO overseeing ground access to that airport, have jurisdiction. This points to a need for airport operators and MPOs to collaborate with their neighbors in long-distance transportation planning, towards a long-distance transportation landscape that works well for the region in terms of both airport ground access as well as air services. Overall, the results can be used towards a more empirically-sound understanding of airport passenger catchments across the region.
This study could be improved upon in several directions. First, similar analyses of more airports in the US Midwest and other regions can verify whether the patterns observed for the three airport categories used in this study hold, or other airport segmentation approaches should be tested. Second, model estimates could also be applied to individual destinations to inform airports and airlines on how their service changes might affect their customer base. Third, controlling for more airport and passenger-specific attributes could help provide further insights into the drivers of long-distance airport choice characteristics. Finally, investigating the choice characteristics of business travellers at similar regional scales would be valuable, as these travelers generate a considerable proportion of airlines’ revenues.

8. Acknowledgements

This study was supported by the Natural Sciences and Engineering Research Council Discovery Grants program. The authors thank Dr. Megan Ryerson at the University of Pennsylvania for the Market Locator dataset acquisition. The authors also thank Patricia Rowe of Milwaukee Mitchell International Airport for providing valuable insights and discussions.

9. References


https://www.scopus.com/record/display.uri?eid=2-s2.0-84905480755&origin=inward&txGid=dabfa4a0a039d431270732cffe55637.

https://www.faa.gov/airports/planning_capacity/passenger_allcargo_stats/passenger/


https://doi.org/10.1016/j.jairtraman.2010.10.010.

https://doi.org/10.1016/j.jtrangeo.2006.08.001.

https://doi.org/10.1080/17445647.2016.1152917.

https://doi.org/10.1016/0965-8564(94)90016-7.

https://doi.org/10.1016/j.jtrangeo.2020.10.014.


https://doi.org/10.3141/2042-01.


Appendix A Final Destination Airports Considered
### Appendix B MMNL ASC Estimates for Short, Medium and Long Routes

<table>
<thead>
<tr>
<th>Airport dummy</th>
<th>Moment</th>
<th>Short route</th>
<th>Medium route</th>
<th>Long route</th>
</tr>
</thead>
<tbody>
<tr>
<td>( I(\text{ATW} = 1) ) (base)</td>
<td>Mean</td>
<td>0 (-)</td>
<td>0 (-)</td>
<td>0 (-)</td>
</tr>
<tr>
<td></td>
<td>Std. dev</td>
<td>0 (-)</td>
<td>0 (-)</td>
<td>0 (-)</td>
</tr>
<tr>
<td>( I(\text{AZO} = 1) )</td>
<td>Mean</td>
<td>-1.718** (0.884)</td>
<td>-1.928 (0.440)</td>
<td>-0.626** (1.758)</td>
</tr>
<tr>
<td></td>
<td>Std. dev</td>
<td>0.147** (1.314)</td>
<td>1.640** (0.883)</td>
<td>0.811 (1.526)</td>
</tr>
<tr>
<td>( I(\text{BMI} = 1) )</td>
<td>Mean</td>
<td>0.001** (0.721)</td>
<td>2.049 (0.435)</td>
<td>8.761 (1.974)</td>
</tr>
<tr>
<td></td>
<td>Std. dev</td>
<td>3.370 (0.624)</td>
<td>3.230 (0.367)</td>
<td>5.698 (1.863)</td>
</tr>
<tr>
<td>( I(\text{CMI} = 1) )</td>
<td>Mean</td>
<td>-3.946 (1.102)</td>
<td>-1.480 (0.531)</td>
<td>2.430** (1.508)</td>
</tr>
<tr>
<td></td>
<td>Std. dev</td>
<td>3.948 (0.796)</td>
<td>1.352** (0.760)</td>
<td>4.965 (1.863)</td>
</tr>
<tr>
<td>( I(\text{CWA} = 1) )</td>
<td>Mean</td>
<td>-2.508 (0.774)</td>
<td>-2.799 (0.470)</td>
<td>-5.264 (1.907)</td>
</tr>
<tr>
<td></td>
<td>Std. dev</td>
<td>1.703** (1.230)</td>
<td>0.975 n/s (0.750)</td>
<td>0.417 (2.299)</td>
</tr>
<tr>
<td>( I(\text{DBQ} = 1) )</td>
<td>Mean</td>
<td>-5.670 (1.117)</td>
<td>-6.985 (1.068)</td>
<td>-0.919** (1.911)</td>
</tr>
<tr>
<td></td>
<td>Std. dev</td>
<td>0.981** (1.124)</td>
<td>1.919** (2.892)</td>
<td>3.495 (1.609)</td>
</tr>
<tr>
<td>( I(\text{DSM} = 1) )</td>
<td>Mean</td>
<td>3.737 (0.994)</td>
<td>6.525 (0.587)</td>
<td>8.642* (4.116)</td>
</tr>
<tr>
<td></td>
<td>Std. dev</td>
<td>3.301 (0.568)</td>
<td>6.285 (0.647)</td>
<td>8.972 (3.383)</td>
</tr>
<tr>
<td>( I(\text{DTW} = 1) )</td>
<td>Mean</td>
<td>7.677 (0.922)</td>
<td>7.167 (0.547)</td>
<td>16.509 (3.291)</td>
</tr>
<tr>
<td></td>
<td>Std. dev</td>
<td>2.023 (0.613)</td>
<td>0.648** (0.536)</td>
<td>0.384 (0.897)</td>
</tr>
<tr>
<td>( I(\text{FWA} = 1) )</td>
<td>Mean</td>
<td>2.161 (0.694)</td>
<td>1.068* (0.427)</td>
<td>2.082** (1.677)</td>
</tr>
<tr>
<td></td>
<td>Std. dev</td>
<td>4.124 (0.808)</td>
<td>1.862 (0.600)</td>
<td>5.343 (1.587)</td>
</tr>
<tr>
<td>( I(\text{GRB} = 1) )</td>
<td>Mean</td>
<td>-0.215** (0.346)</td>
<td>-0.314** (0.338)</td>
<td>0.465** (0.729)</td>
</tr>
<tr>
<td></td>
<td>Std. dev</td>
<td>1.198** (1.051)</td>
<td>0.997* (0.428)</td>
<td>0.154 (1.164)</td>
</tr>
<tr>
<td>( I(\text{GRR} = 1) )</td>
<td>Mean</td>
<td>3.523 (0.727)</td>
<td>3.586 (0.418)</td>
<td>7.571 (2.178)</td>
</tr>
<tr>
<td></td>
<td>Std. dev</td>
<td>2.563 (0.680)</td>
<td>3.087 (0.380)</td>
<td>4.825 (1.262)</td>
</tr>
<tr>
<td>( I(\text{IND} = 1) )</td>
<td>Mean</td>
<td>8.210 (0.878)</td>
<td>8.241 (0.600)</td>
<td>17.587 (3.367)</td>
</tr>
<tr>
<td></td>
<td>Std. dev</td>
<td>0.376** (0.340)</td>
<td>0.156** (0.395)</td>
<td>4.382 (1.053)</td>
</tr>
<tr>
<td>( I(\text{MKE} = 1) )</td>
<td>Mean</td>
<td>5.993 (0.686)</td>
<td>7.542 (0.482)</td>
<td>13.447 (2.560)</td>
</tr>
<tr>
<td></td>
<td>Std. dev</td>
<td>0.358** (0.497)</td>
<td>4.161 (0.389)</td>
<td>2.751 (1.764)</td>
</tr>
<tr>
<td>( I(\text{MLI} = 1) )</td>
<td>Mean</td>
<td>0.939** (0.944)</td>
<td>0.494** (0.401)</td>
<td>7.951 (2.136)</td>
</tr>
<tr>
<td></td>
<td>Std. dev</td>
<td>1.241** (1.589)</td>
<td>2.050 (0.490)</td>
<td>0.339 (2.190)</td>
</tr>
<tr>
<td>( I(\text{MSN} = 1) )</td>
<td>Mean</td>
<td>4.189 (0.632)</td>
<td>4.399 (0.377)</td>
<td>3.202* (1.440)</td>
</tr>
<tr>
<td></td>
<td>Std. dev</td>
<td>0.700** (0.700)</td>
<td>2.547 (0.411)</td>
<td>3.321 (1.559)</td>
</tr>
<tr>
<td>( I(\text{MSP} = 1) )</td>
<td>Mean</td>
<td>9.237 (1.044)</td>
<td>9.707 (0.643)</td>
<td>-12.067** (2.718)</td>
</tr>
<tr>
<td></td>
<td>Std. dev</td>
<td>6.275 (1.491)</td>
<td>0.228** (0.468)</td>
<td>14.408 (3.973)</td>
</tr>
<tr>
<td>( I(\text{ORD} = 1) )</td>
<td>Mean</td>
<td>9.616 (1.018)</td>
<td>10.163 (0.670)</td>
<td>12.508 (2.720)</td>
</tr>
<tr>
<td></td>
<td>Std. dev</td>
<td>3.716 (0.584)</td>
<td>3.559 (0.392)</td>
<td>0.059 (1.590)</td>
</tr>
<tr>
<td>Airport dummy</td>
<td>Moment</td>
<td>Short route</td>
<td>Medium route</td>
<td>Long route</td>
</tr>
<tr>
<td>---------------</td>
<td>--------</td>
<td>-------------</td>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>1.651* (0.784)</td>
<td>3.563 (0.581)</td>
<td>13.504 (3.032)</td>
</tr>
<tr>
<td></td>
<td>Std. dev</td>
<td>1.736* (1.020)</td>
<td>0.705** (0.887)</td>
<td>4.547 (2.204)</td>
</tr>
<tr>
<td>I(PIA = 1)</td>
<td>Mean</td>
<td>2.542 (0.762)</td>
<td>1.202* (0.593)</td>
<td>3.133** (1.898)</td>
</tr>
<tr>
<td></td>
<td>Std. dev</td>
<td>2.092* (0.823)</td>
<td>1.640* (0.747)</td>
<td>2.518 (1.671)</td>
</tr>
<tr>
<td>I(SBN = 1)</td>
<td>Mean</td>
<td>-2.192** (1.281)</td>
<td>0.904** (0.720)</td>
<td>6.552 (2.316)</td>
</tr>
<tr>
<td></td>
<td>Std. dev</td>
<td>0.022** (2.792)</td>
<td>6.310* (2.479)</td>
<td>6.550 (4.496)</td>
</tr>
<tr>
<td>I(SPI = 1)</td>
<td>Mean</td>
<td>2.671* (1.081)</td>
<td>4.494 (0.892)</td>
<td>10.273 (2.927)</td>
</tr>
<tr>
<td></td>
<td>Std. dev</td>
<td>6.360 (1.078)</td>
<td>4.829 (0.797)</td>
<td>6.892 (2.075)</td>
</tr>
</tbody>
</table>

1 * significant at the 95% confidence level.
2 ** significant at the 90% confidence level.
3 n/s not significant at the 90% confidence level.
4 All other parameters are significant at 99% confidence level.