A Parameterized Consideration Set Model for Airport Choice:

An Application to the San Francisco Bay Area

Gözen Başar and Chandra Bhat

Department of Civil Engineering, Bldg ECJ, Suite 6.8

The University of Texas at Austin, Austin, TX, 78712

Phone: 512-471-4535, Fax: 512-475-8744,

Email: gozen@mail.utexas.edu, bhat@mail.utexas.edu

ABSTRACT

Airport choice is an important air travel-related decision in multiple airport regions. This paper proposes the use of a probabilistic choice set multinomial logit (PCMNL) model for airport choice that generalizes the multinomial logit model used in all earlier airport choice studies. The paper discusses the properties of the PCMNL model, and applies it to examine airport choice of business travelers residing in the San Francisco Bay Area. Substantive policy implications of the results are discussed. Overall, the results indicate that it is important to analyze the choice (consideration) set formation of travelers. Failure to recognize consideration effects of air travelers can lead to biased model parameters, misleading evaluation of the effects of policy action, and a diminished data fit.

1. INTRODUCTION

Intercity travel has grown steadily over the past decade, and recent studies suggest that such travel is only likely to grow even further through the next two decades. Within the context of intercity travel, air travel is the fastest growing travel mode in the United States. Notwithstanding the events of September 11, 2001, projections suggest that the number of air travelers in the U.S. will double in this first decade of the 21st century. Further, airports are increasingly serving as freight gateways to facilitate long-distance commodity movement nationally and internationally. As the number of air travelers and amount of air freight movements increase, so will the contribution of airport-related travel to overall urban traffic levels. In addition, increases in person travel and freight lead to higher staffing needs at airports, thus increasing commuting travel to/from airports.

In contrast to the increasing contribution of air travel to urban travel, airport-related travel is still treated in a rather coarse and simplified manner within the urban travel modeling framework of most Metropolitan Planning Organizations in the State and the Country. In particular, airports are identified as "special attractors" and assigned a certain number of trip attractions, without adequate systematic analysis of the spatial and temporal patterns of the trip attractions. It is important for transportation agencies to consider a more systematic approach to analyze and forecast airport-related personal travel, so that improved predictions of traffic characteristics and traffic levels on urban roadways may be achieved. A systematic analysis of airport travel is also important for mobile-source emissions forecasting.

There are several dimensions characterizing air traveler decisions that impact the spatial and temporal distribution of trips to the airport. For residents of an urban area, some of the first decisions regarding inter-urban travel may include whether to travel away from the urban area and to where, the duration of the trip, and the mode for the inter-urban trip (*i.e.*, whether to travel by air, or some other mode). If air is the mode of choice, the relevant decisions include the destination airport, the origin airport in a multi-airport urban area, the desired arrival time at the destination (which impacts the desired flight departure time at the origin), the originating location and departure time of the ground access trip to the origin airport, and the access mode of transport to the airport. In addition to these choices, other air traveler decisions that would be of relevance to air carriers and airport management include air carrier choice, fare class of travel, and method of purchase of tickets (see Harvey, 1987 for a discussion of these air travel-related choice dimensions).

The many dimensions of air travel identified above are clearly inter-related. Ideally, the analyst would prefer a modeling structure that models all these dimensions jointly. But such a joint framework is infeasible in practice, and thus the analyst needs to assume a sequential structure that may be assumed to reasonably represent the air travel choice process. Harvey (1987) provides one such possible hierarchical sequence.

An important choice dimension, which precedes most other air travel decisions in Harvey's framework, is the origin departure airport choice in a multi-airport urban region. A good understanding of the factors underlying passenger's origin airport choice in multi-airport urban regions can enable airport management and airline carriers to attract passengers, upgrade airport facilities and equipment to meet projected air travel demands, and determine airport staffing needs. It can also aid Metropolitan Planning Organizations in forecasting travel demand in the urban region, and in planning transportation networks to/from airports.

Several earlier studies have examined airport choice in a multi-airport region. Some of these studies have focused on airport choice in isolation (see Skinner, 1976; Harvey, 1987;

Ashford and Benchemam, 1987; Ozoka and Ashford, 1989; Innes and Doucet, 1990; Thompson and Caves, 1993; and Windle and Dresner, 1995), while others have examined airport choice along with other dimensions of air travel (see Ndoh *et al.*, 1990; Furuichi and Koppelman, 1994; Pels *et al.*, 2001; and Pels *et al.*, 2003). These earlier studies have focused on different urban areas and, sometimes, different population groups (such as business travelers versus leisure travelers and residents versus non-residents). However, a common finding in all these studies is that access time to the airport and frequency of service from the airport to the desired destination are the dominant factors affecting airport choice. Several of these studies also suggest that a simple measure of access time to the airport; *i.e.*, auto access time; performs as well as more complex formulations that consider multiple modes and both access time and access cost. In addition, many earlier studies find that airfare is not a significant factor in airport choice for business travelers, though a few studies find airfare to affect airport choice for non-business travelers[1].

The current study contributes to the existing body of literature by focusing on airport choice in the San Francisco Bay Urban Area context. An important characteristic of the current study is its recognition that travelers may not consider all the available airports when making the choice of their departure airport. Earlier research on choice set generation has indicated the important impact of consideration effects on consumer choice (see, for example, Roberts and Lattin, 1991; Ben-Akiva and Boccara, 1995; Chiang *et al.*, 1999). However, all the airport choice models discussed earlier assume that each traveler makes a choice from the full set of available airports, where an airport is assumed to be available if there is at least one flight (direct or connecting) from the airport to the destination city. Such an assumption is rather untenable

[1] A comprehensive review of previous airport choice studies has been completed recently by the authors, and a review report is available from the authors.

because an individual's choice set is likely to depend on the traveler's specific sociodemographic, informational, psychological, and societal contexts as well as subjective criteria associated with individual attitudes/perceptions. For example, an individual may consider a particular airport to be too far away to be even considered, while another individual may consider this distance to be acceptable. Similarly, an individual may eliminate from consideration any airport that does not have airline club lounges, while another may include airports without airline club lounges in her/his choice set. Thus, it is important to recognize that different travelers may, and in general will, consider different sets of alternatives.

To be sure, considering the choice set formation process along with the actual choice process is not merely an esoteric econometric issue. Earlier research in the transportation and marketing fields has indicated that failure to properly specify the choice set considered by consumers can lead to biased choice model parameters, a lack of robustness in parameter estimates, and violations of the independence from irrelevant alternatives assumption (see Shocker *et al.*, 1991; Swait, 1984; and Williams and Ortuzar, 1982). On the other hand, the explicit incorporation of consideration effects has both methodological and managerial benefits. Methodologically, the incorporation of consideration effects can lead to a more accurate prediction of the choice process being modeled (see Gensch, 1987; Chiang *et al.*, 1999; and Swait, 2001). Such prediction gains will result in improved forecasting of travel demand to/from airports. Managerially, the recognition of consideration effects can help determine the relative effects of policy relevant variables on consideration and choice, and thus aid in a comprehensive understanding of the impacts of policy actions (as we discuss in sections 4 and 5). The important point to note here is that regardless of the relative utility of an airport compared to other airports in a traveler's choice set, the airport will not be chosen if it is not first considered (see Andrews and Srinivasan, 1995).

In addition to the methodological issue of modeling the choice set generation process and airport choice from the choice set, the current paper also considers the impact of sociodemographic and trip characteristics of the traveler on airport choice. Harvey (1987) is one of the only earlier studies that recognizes demographic impacts, but that study did not find any statistically significant effects of personal characteristics on airport choice.

The rest of this paper is structured as follows. The next section discusses the model structure. Section 3 presents the data source and sample formation procedures. Section 4 describes empirical results. The final section highlights the important findings for the paper.

2. MODEL STRUCTURE

2.1 Background

The model structure used in this paper is based on Manski's (1977) original two-stage choice paradigm, which includes a probabilistic choice set generation model in the first stage followed by the choice of airport from a given choice set.

The first stage uses a probabilistic choice set generation mechanism because the actual choice set of travelers is unobserved to the analyst and, therefore, cannot be determined with certainty by the analyst. Within the class of probabilistic choice set generation models, we adopt Swait and Ben-Akiva's (1987) random constraint-based approach to choice formation (for a detailed discussion of other approaches to probabilistic choice set generation, see Ben-Akiva and Boccara, 1995). In the random constraint-based approach, an airport is excluded from the choice set if the consideration utility for that airport is lower than some threshold consideration utility level (the reader will note that the consideration of an airport is determined only by the threshold level of that airport, not by any comparisons to the thresholds of other airports). Since the threshold utility level is not observed to the analyst, the exclusion of an airport from the choice set becomes probabilistic. In the current study, we allow the consideration utility to vary across individuals, so that the consideration probability of each airport varies across individuals. Almost all earlier applications of probabilistic choice set generation have used the same consideration probability across individuals (but see Andrews and Srinivasan, 1995 and Swait, 2001)[2].

The second stage airport choice model, given the choice set, is based on the familiar multinomial logit formulation. At this stage, the utilities of the airports in the choice set are compared directly with each other in a utility maximizing process. The difference in the process at the choice set generation and choice stages enables a change in an attribute associated with an airport to have two separate effects: a consideration effect (*i.e.*, the impact on the consideration set of airports) and a choice effect (*i.e.*, the impact on the choice of an airport, given that the airport is considered by the individual).

2.2 Formulation

The model formulation in this section is developed assuming that all airports are feasible for each traveler (though not all of the airports may be considered by each traveler). This assumption simplifies the presentation and is consistent with the empirical context of the current paper,

[2] Note that the probability of a choice set c in the current paper is obtained by first modeling the consideration probabilities for each alternative individually and multiplying the individual consideration probabilities appropriately. Such a procedure assumes that the consideration probabilities for any two alternatives are independent, except for the correlations due to common observed factors affecting consideration probabilities. Another approach to model the choice set is to directly consider each possible nonempty subset of the full choice set as an alternative. However, this latter approach leads to an explosion in the number of choice set alternatives even for small numbers of alternatives in the full choice set.

where each airport has at least one direct or connecting flight in the day to each traveler's destination airport.

Let the consideration utility of airport i (i=1,2,…,I) for individual q be U_{qi}. The alternative is included in the choice set if this consideration utility exceeds a certain threshold and is eliminated if not. Since the threshold is not observed to the analyst, it is considered as a random variable. In the current paper, we assume this random threshold to be standard logistically distributed (Swait, 2001 also makes the same assumption). Then, the probability that alternative i is considered by individual q can be written as:

$$M_{qi} = \frac{1}{1+e^{-\gamma w_{qi}}}, \tag{1}$$

where w_{qi} is a column vector of observed attributes for individual q and alternative i (including a constant) and γ is a corresponding column vector of coefficients to be estimated (this coefficient provides the impact of attributes on the consideration probability of alternative i).

Next, assume that the randomly-distributed threshold for each alternative is independent of the threshold values of other alternatives. The overall probability of a choice set c for individual q may then be written as:

$$P_q(c) = \frac{\prod_{i \in c} M_{qi} \prod_{j \notin c} (1 - M_{qj})}{1 - \prod_{i=1}^{I} (1 - M_{qi})}, \tag{2}$$

where the denominator is a normalization to remove the choice set with no alternatives in it.

The choice of airport from a given choice set can be written, using a multinomial logit formulation, as:

$$P_{qi} \mid c = \frac{e^{\beta' x_{qi}}}{\sum_{j \in c} e^{\beta' x_{qj}}} \text{ if } i \in c \qquad (3)$$
$$= 0 \text{ if } i \notin c \,,$$

where x_{qi} is a column vector of exogenous variables and β is a column vector of coefficients indicating the effect of variables at the choice stage.

Finally, the unconditional probability of choice of alternative *i* can be written as follows:

$$P_{qi} = \sum_{c \in G} (P_{qi} \mid c) \cdot P_q(c) \;, \qquad (4)$$

where *G* is the set of all nonempty subsets of the master choice set of all airport alternatives. The membership of *G* will include ($2^I - 1$) elements. For example, in a three airport case, denoted as {A,B,C}, *G* includes the following choice sets: {A}, {B}, {C}, {A,B}, {B,C}, {A,C}, {A,B,C}.

The log-likelihood function for the estimation of the parameters β and γ is:

$$\mathrm{L}\ (\beta, \gamma) = \sum_q \sum_i y_{qi} \cdot \log P_{qi}(\beta, \gamma) \,, \qquad (5)$$

where y_{qi} is a dummy variable taking the value 1 if individual *q* chooses airport *i* and 0 otherwise. Maximization of the log-likelihood function is accomplished using the GAUSS matrix programming language.

2.3 Properties

The parameterized probabilistic choice set multinomial logit (PCMNL) model structure presented in the previous section nests the multinomial logit structure as a special case. In particular, the probability function of Equation (4) collapses to the MNL model if $M_{qi} = 1$ for all alternatives *i* and all individuals *q* (also note that $M_{qi} \to 1$ when $\gamma' w_{qi} \to +\infty$ for all *i* and *q*). In this situation, $P_q(c)$ = 0 for all choice sets *c* that are subsets of the master choice set and $P_q(c)$ = 1 for the master choice set, which is equivalent to assuming that all individuals consider all airports.

The disaggregate-level elasticity effects in the PCMNL model can be computed from the probability expression in Equation (4) in a straightforward manner (however, we are not aware of any earlier study presenting these expressions). In the following presentation of elasticity expressions, we suppress the index q for individuals for notation ease. Let δ_i^c be a dummy variable taking the value 1 if choice set c contains airport i and 0 otherwise, and let δ_{ij}^c be another dummy variable taking the value 1 if choice set c contains both airports i and j and 0 otherwise. Also, define B_i as follows, where B_i represents the probability that the individual's choice set includes alternative i:

$$B_i = \sum_{c \in G} \delta_i^c P(c) = \frac{M_i}{1 - \prod_k (1 - M_k)} . \tag{6}$$

Then the self- and cross-elasticities of a change in the m^{th} attribute of an airport $i\,(z_{im})$ that appears at both the consideration stage and choice stage can be written as follows:

$$\begin{aligned} \eta_{z_{im}}^{P_i} &= \left[(1 - B_i)\gamma_m + \frac{1}{P_i} \sum_{c \in G} \left\{ (P_i \mid c)(1 - P_i \mid c) P(c) \beta_m \right\} \right] z_{im} \\ \eta_{z_{im}}^{P_j} &= \left[\left\{ \frac{1}{P_j} \sum_{c \in G} (P_j \mid c) P(c) \cdot \delta_{ij}^c - B_i \right\} \gamma_m + \frac{1}{P_j} \sum_{c \in G} \left\{ (-P_i \mid c)(P_j \mid c) P(c) \beta_m \right\} \right] z_{im} \end{aligned} \tag{7}$$

Each of the expressions above comprise two terms. The first term represents the consideration elasticity and captures the impact of a change in z_{im} on the consideration of airport i in the self-elasticity expression and on the consideration of airport j relative to airport i in the cross-elasticity expression. The second term represents the substitution elasticity at the choice stage conditional on the alternative being available in the choice set. Note that for a variable that does not appear in the consideration stage, only the substitution elasticity applies in each of the expressions. On the other hand, for a variable that does not appear at the choice stage, only the consideration elasticity applies. In any case, the cross-elasticity expression is a function of the choice probability for alternative j. Thus, the PCMNL model does not exhibit the IIA property of the MNL model. It is also easy to verify that the self- and cross-elasticity expressions collapse to those of the MNL when all airports are considered.

3. DATA SOURCE AND SAMPLE FORMULATION

The primary data source for this study is an air passenger survey conducted by the Metropolitan Transportation Commission in the San Francisco Bay Area. This survey was administered to randomly selected travelers in August and October of 1995 at four airports: San Francisco International (SFO), San Jose International (SJC), Oakland International (OAK), and Sonoma County (STS). Information collected in the survey included purpose of travel, destination, number in the traveling party, mode of transport to the airport, airline carrier, and flight details. In addition, sociodemographic attributes of the traveler were also obtained.

In the current research, the survey responses from the three major Bay Area airports; SFO, SJC, and OAK; are used because of the very low share of travelers using the Sonoma County airport. For ease in data preparation and assembly, the top thirty destinations from these three Bay Area airports are identified from the sample and the airport choice of Bay Area residents to these top destinations are considered. These top thirty destinations are served from each of the three Bay Area airports, either through direct flights and/or connecting flights. Thus, all the three airports are available as potential choices, though not all of them may be considered by travelers.

The air travel market is segmented, for the purpose of our analysis, into business and non-business trip purposes. To narrow the focus, we consider only business trips in this paper. The final business sample comprises 1,918 observations, of which 1,618 observations are used for estimation and the remaining 300 observations are set aside as a validation sample for evaluating the performance of an ordinary multinomial logit (MNL) model and the parameterized probabilistic choice set multinomial logit (PCMNL) model of this paper. The sample shares and the market shares in the estimation sample are presented in Table 1. As can be observed, there is an oversampling of travelers flying out of San Jose in the airport survey (the actual shares of airport choice in the population are obtained from the Bureau of Transportation Statistics). Since the sample is choice-based with known aggregate shares, we employ the Weighted Exogenous Sample Maximum Likelihood (WESML) method proposed by Manski and Lerman (1977) in estimation. This method weights the log-likelihood value for each individual in Equation (5) by the ratio of the market share of the airport chosen by the individual to the sample share of the airport chosen by the individual (the resulting weights are presented in the final column of Table 1). Maximizing the resulting likelihood function provides consistent estimates of the parameters. The asymptotic covariance matrix of parameters is computed as $H^{-1}\Delta H^{-1}$, where H is the hessian and Δ is the cross-product matrix of the gradients (H and Δ are evaluated at the estimated parameter values). This provides consistent standard errors of the parameters (Börsch-Supan, 1987).

In addition to the air passenger travel survey, three other secondary data sources are used to develop the final sample. The first is a zone-to-zone ground access level of service file, obtained from the Metropolitan Transportation Commission in Oakland. This information is appropriately appended to the sample observations based on the originating zone of departure to the airport and the zone that contains each airport. In the current analysis, we use the level-of-service (time and cost) values corresponding to the highway mode, since a majority of the trips to the airport are pursued by a private or rental car. The second secondary data source used in the analysis is the daily flight frequency from each Bay Area airport to the thirty destination airports, obtained from the 1995 Official Airline Guide (Official Airline Guide Market Analysis, 1995)[3]. This information is appended to the sample observations based on the origin-destination airport pair and the day of week of travel. The third source is on-time flight statistics for nonstop flights from each airport to each destination, obtained from the Bureau of Transportation Statistics (BTS). These data provide the percentage of late flights, defined as the percentage of flights delayed beyond 15 minutes of the scheduled departure time (The BTS on-time flight statistics are for 1997, and its use in the current analysis assumes the absence of significant changes between 1995 and 1997).

The three secondary data sources discussed above provide measures of the quality of service offered by each airport for the traveler's trip.

4. EMPIRICAL ANALYSIS

4.1 Variable Specification

The choice of variables for potential inclusion was guided by previous empirical work on airport choice modeling, intuitive arguments regarding the effects of exogenous variables, and data availability considerations. We considered three broad classes of variables for inclusion: (1) quality of service variables, (2) interactions of sociodemographics with quality of service, and (3) interactions of trip characteristics with quality of service.

[3] These include nonstop flights and flights with a stop but no change in equipment.

The quality of service variables, as discussed earlier, included ground-access level of service variables (time and cost) and air travel level-of-service variables (flight frequency to destination and percentage of late flights). Traveler sociodemographics considered in the analysis included the sex, age, and household income of the traveler. Finally, the trip characteristics explored in the specifications included the following dummy variables: (a) an “alone” variable identifying whether or not the individual was traveling alone, (b) a “short trip” variable representing if the traveler was away for fewer than 2 nights or 2 or more nights, (c) a “car used to reach airport” variable indicating whether the traveler used a car (private or rented) to reach the airport, (d) a “weekday” variable indicating if the trip was pursued on a weekday or the weekend, and (e) a “left to airport from work” variable identifying if the traveler left to the airport from work or from a nonwork location.

Several nonlinear forms for capturing the effect of access time and flight frequency were explored in our analysis. But the simple linear functional form for access time and flight frequency performed as well as the more complex functional forms. We arrived at the final specification based on a systematic process of eliminating variables found to be insignificant in previous specifications and based on considerations of parsimony in representation.

4.2 Estimation Results

The results of the multinomial logit (MNL) model and the parameterized probabilistic choice set multinomial logit (PCMNL) model are presented in Table 2 and discussed in the subsequent two sections.

4.2.1 The MNL Model Results

The coefficients on the access time variable in the multinomial logit model indicate, as one would expect, that business travelers are averse to traveling long durations to reach an airport. This is particularly the case for individuals traveling alone and women travelers. The coefficients on the frequency variable indicate a preference for airports that have frequent flight service to the traveler's destination. Individuals traveling alone, in particular, place a premium on frequency. This result, along with the higher access time sensitivity of individuals traveling alone, suggests that time is less onerous when traveling in a group (perhaps because of the opportunity to socialize or conduct business when traveling together). The results also indicate the lower sensitivity of women and high-income individuals to frequency of service. The latter result is a little surprising, but may be a reflection of high-income individuals traveling at narrow peak-period time windows of the day, and thus not being sensitive to the frequency of flights over the entire day. Frequency of service does not impact airport choice for high-income women travelers.

4.2.2 The PCMNL Model Results

The PCMNL model includes estimates of the probabilistic choice set generation model as well as the airport choice model. The coefficients at the consideration stage provide estimates of the γ vector in Equation (1). Table 2 shows that the coefficients on the access time and frequency variables at the consideration stage are statistically significant, indicating variation in the consideration of each airport across individuals. In particular, airports that are farther away and/or that have a low frequency of flights are less likely to be considered by individuals. As one would expect, these effects are magnified on weekdays compared to weekends. Additionally, women appear to be more willing than men to consider airports that are distant from their point of departure to the airport. This gender difference at the consideration stage may be a manifestation of the disparity in decision mechanisms between men and women; women appear to be more inclusive at the consideration stage, while men focus on quickly narrowing down options to consider.

The coefficient estimates in the choice stage in the PCMNL model have interpretations that are similar to those in the MNL model. However, there are differences in the magnitude of the access time impacts. Specifically, the access time effects at the choice stage are higher than the corresponding MNL estimates. The reason is that airports that are very far away are "removed" from consideration in the PCMNL model. For example, consider an individual with one close airport and two very distant airports, and assume that this individual considers only the close airport. For this individual, access time has no impact (by definition) at the choice stage (the probability of choice of the close airport is one, given that the choice set includes only that airport). Thus, the sensitivity to access time at the choice stage in the PCMNL model is automatically based on data from individuals who have a high probability of consideration of two or more airports, and who are sensitive to access time at the choice stage. The MNL model, on the other hand, includes relatively "captive" individuals in the choice model estimation, despite these individuals not being sensitive to access time. The result is a dilution of the sensitivity to access time in the MNL choice model. The impact of frequency at the choice stage of the PCMNL model is not very different from the MNL model.

The combination of results at the consideration and choice stages shows that access time is less important for women when developing the perception "space" of availability of airports, but is more important for women when choosing an airport from the choice set of available airports.

4.3 Trade-off Between Access Time and Frequency of Service

The coefficients on time and frequency can be used to examine the trade-offs between the two determinants of airport choice. For example, the MNL model indicates that male, low-income, individuals traveling in a group would be willing to travel about 6 minutes [=0.411/(6.964/100)] longer if the frequency of flight service were to be increased by ten flights per day. The corresponding values for other traveler subgroups are provided in Table 3 for both the MNL and PCMNL models. In general, these results indicate that access time is the dominant determinant of airport choice for business travelers, particularly for high-income group travelers. In addition, the PCMNL values indicate that, at the choice stage, access time is an even more dominant determinant than suggested by the MNL model.

The time values of frequency can also be computed for the consideration stage from the PCMNL model. Interestingly, these values are very high. An additional flight per day from an airport has the same impact on consideration utility as 18 less minutes to that airport for male weekend travelers, 90 less minutes for female weekend travelers, 9.5 less minutes for male weekday travelers, and 13.5 less minutes for female weekday travelers. These results show the relatively dominant effect of frequency at the consideration stage, especially on weekends.

4.4 Substantive Policy Implications

The relative effects discussed above provide useful information about the effects of access time and frequency on choice in the MNL model, and separately on consideration and choice in the

PCMNL model. However, these effects do not provide a measure of the absolute magnitude of impacts. Further, in the PCMNL model, the overall effects of access time and frequency are not directly discernible from the coefficients at the consideration and choice stages.

To examine the overall effects of access time and frequency, we now compute the aggregate self- and cross-elasticities. These aggregate elasticities provide the proportional change in the expected market shares of each airport in response to a uniform percentage improvement in access time and frequency across all individuals. The aggregate self- and cross-elasticities can be obtained from the disaggregate-level elasticities presented in Equation (7).

Table 4 shows the elasticity effects for the MNL and PCMNL models. Several common conclusions may be drawn from the elasticities of the MNL and PCMNL models. First, in the overall, access time is a more important determinant of airport choice than is air service frequency. This is consistent with several earlier studies on airport choice. Second, the self-elasticities indicate that Oakland International is best positioned to improve its market share through improvements in its quality of service (note the higher self-elasticity effects for Oakland compared to the self-elasticity effects of the other two airports). Third, San Francisco International has tremendous "clout" in the market, since it can easily negate attempts by other airports to draw away share by making its own marginal service improvements (see the much higher cross-elasticities corresponding to improvements in SFO's quality of service compared to the cross-elasticities corresponding to improvement in the quality of service of other airports).

The substantive policy implications from the MNL and PCMNL models, while similar in some ways, are also quite different in others. First, compared to the MNL model, the PCMNL model indicates substantially lower self- and cross-elasticities corresponding to access time. If the PCMNL model is a more appropriate model (as we will clearly demonstrate in the next section), use of the MNL model would overestimate the potential gain in an airport's market share due to an improvement in access time to that airport and would overestimate the reduction in market share of other airports due to such an access time improvement. Second, the PCMNL model shows higher self- and cross-elasticities corresponding to improvement in air frequency from San Jose and Oakland airports. This can be attributed to the strong impact of air frequency on consideration of an airport in the PCMNL model, as discussed in the previous section. The reason why such an effect does not extend to San Francisco is that San Francisco already has a very high consideration level in the market. In fact, the overall consideration level can be estimated from the parameter estimates in Table 2. Defining S_i as the share of individuals who consider airport *i* when making a choice, we can write:

$$S_i = \frac{\sum_q w_q \sum_{c \in G} \delta_i^c P_q(c)}{Q}, \tag{8}$$

where w_q is the weight for individual *q*, *Q* is the total number of individuals in the sample, and other quantities are as defined earlier. The estimated values of airport consideration are 99.4% for SFO, 77.2% for SJC, and 70.7% for OAK. Clearly, there is little room to increase the consideration level of SFO, which is the reason for the low self- and cross-elasticities corresponding to air service frequency improvement for SFO.

To summarize, the substantive implications for policy analysis from the MNL and PCMNL models are different in the current empirical context. These differences suggest the need to apply formal statistical tests to determine the structure that is most consistent with the data. This is the focus of the next section.

4.5 Measures of Data Fit

We evaluate the fit of the MNL and PCMNL models in both the estimation sample and a validation sample. In the estimation sample, we compute the standard measures of fit, including the log-likelihood at convergence and the adjusted likelihood ratio index. The adjusted likelihood ratio index is defined with respect to the log-likelihood at market shares:

$$\bar{\rho}^2 = 1 - \frac{\mathcal{L}(\hat{\beta}) - M}{\mathcal{L}(c)}, \tag{9}$$

where $\mathcal{L}(\hat{\beta})$ and $\mathcal{L}(c)$ are the log-likelihood functions at convergence and at market shares, respectively, and M is the number of parameters estimated in the model (besides the alternative specific constants of the choice model). In addition, we also compute the average probability of correct prediction, which is computed as $Q^{-1}\sum_q w_q \sum_i y_{qi}\hat{P}_{qi}$, where $\hat{P}_{qi}$ is the estimated probability of individual q choosing airport i at the convergent values.

The results for the estimation sample are presented in the second main column of Table 5. The adjusted likelihood ratio index and the average probability of correct prediction clearly favor the PCMNL model (see the last two rows of the table). A formal statistical nested likelihood ratio test between the convergent log-likelihood values of the two models indicates a value of 400.0, which is larger than the corresponding chi-squared value with 8 degrees of freedom at any reasonable level of significance.

We also evaluate the performance of the MNL and PCMNL models on a holdout (validation) sample to verify that the results obtained from the estimation sample are not an artifact of overfitting. We set aside 300 observations for validation such that the shares in the validation sample are close to the actual market shares (this allows the direct application of the estimated model results to the validation sample, without the need to adjust the airport-specific constants). Two measures of fit are computed in the validation sample. The first is the predictive adjusted likelihood ratio index, which is computed by calculating the predictive log-likelihood function value at the parameter estimates obtained from estimation. The second is the average probability of correct prediction, also computed at the parameter values obtained from estimation. These disaggregate measures of fit are presented in the last two rows of the third main column in Table 5. As can be observed, there is a drop in the adjusted likelihood ratio index from the estimation sample for both the MNL and PCMNL models. But the PCMNL model still provides a value that is higher than the MNL model. The average probability of correct prediction in the validation sample also reflects this superior fit of the PCMNL model. In summary, the PCMNL clearly outperforms the MNL model from a statistical standpoint.

Another more informal, but intuitive, way to compare the two models is to compute the estimated distribution of consideration sets across resident air travelers in the Bay Area. This can be computed as $Q^{-1}\left[\sum_q w_q \hat{P}_q(c)\right]$, where $\hat{P}_q(c)$ is the predicted probability from the PCMNL model of individual q having the consideration set c. The resulting distribution, providing the percentage of individuals with each of the seven possible choice sets, is as follows: SFO only (23.50%), SJC only (0.22%), OAK only (0.12%), SFO and SJC (13.46%), SJC and OAK (0.07%), SFO and OAK (9.83%), and all airports (52.80%). These results indicate that about half of all travelers do not choose from the universal choice set of all the three airports. However, the MNL model assumes that all travelers choose from the universal choice set. Another interesting observation is that about a quarter of all travelers consider only SFO. In summary, these results again highlight the clout of SFO in the consideration perception map of Bay Area air travelers.

5. SUMMARY AND CONCLUSIONS

This paper proposes the use of a probabilistic choice set multinomial logit model (PCMNL) for airport choice analysis that generalizes the commonly used multinomial logit (MNL) model. The PCMNL model takes the form of a random constraint-based approach to choice formation in which an airport is excluded from the choice set if the consideration utility of that airport is lower than a threshold utility level. The choice of airport from a given choice set is based on the usual MNL structure. The properties of the PCMNL model are discussed, including the presentation and interpretation of elasticity expressions.

The PCMNL model is applied to examine the airport choice of business travelers residing in the San Francisco Bay Area. Several important conclusions may be drawn from the empirical analysis. First, as found in earlier studies, access time to the airport and flight frequency are the two primary determinants of airport choice. However, unlike earlier studies, our study indicates variation in sensitivity to these two variables based on traveler demographics and trip characteristics. Specifically, individuals traveling alone and women travelers are more sensitive to access time, and individuals traveling alone are also more sensitive to flight frequency. Further, women and high-income travelers are not very sensitive to flight frequency. In addition, the results from the consideration stage of the PCMNL model indicate that access time and flight frequency affect the consideration of an airport. Second, the access time parameter estimates of the MNL model and the choice stage of the PCMNL model are quite different. This is because the MNL model arbitrarily assumes that all airports are available to all individuals. A comparison of the relative trade-off between access time and frequency from the two models suggests the dominance of access time at the choice stage, particularly in the PCMNL model. However, the PCMNL model also indicates that, in forming perceptions of the availability of airports, flight frequency is the dominating factor. Interestingly, access time is less important to women (relative to men) when forming the perception space of available airports, but is more important to women when choosing an airport from the set of available airports. These results have implications for the design of promotional marketing strategies. For instance, an airport attempting to increase market share by improving access time to its terminals might consider targeting informational campaigns within its traditional catchment area of travelers (*i.e.*, areas in close proximity to the airport) and by targeting women travelers (at airports, or by targeting firms/occupations which are women-dominated). On the other hand, information campaigns regarding frequency improvements are better positioned in areas that are not within the traditional catchment area (*i.e.*, in areas that are distant from the airport) and are likely to be more productive if targeted toward weekend travelers. Clearly, only the PCMNL model is able to offer such comprehensive insights into the effects of variables. Third, the substantive elasticity effects from the MNL and PCMNL models indicate that access time is the most important factor in the choice of an airport. Also, in the San Francisco Bay Area market, San Francisco International has tremendous clout, since it can easily compensate for service improvements at other airports by making marginal improvements in its own service. Between the MNL and the PCMNL model, the PCMNL model predicts a lower overall impact of access time, indicating that the use of the MNL model overestimates the potential gain in airport market share due to an improvement in access time to that airport. On the other hand, the PCMNL model predicts a higher overall impact of flight frequency, suggesting an underestimation of the net gains from improving frequency by the MNL model. Fourth, the PCMNL model clearly outperforms the MNL model in statistical evaluation of data fit in both an estimation sample and a validation sample.

In summary, the application of the PCMNL model to airport choice suggests that it is important to model consideration sets of air travelers. Failure to recognize consideration effects can lead to biased model parameters, misleading evaluations of the effects of policy actions, as well as a considerably diminished data fit.

ACKNOWLEDGEMENTS

The authors would like to thank Ken Vaughn and Chuck Purvis of the Metropolitan Transportation Commissions (MTC) in Oakland for providing help with data related issues. Two anonymous referees provided valuable comments on an earlier version of the paper. The authors are also grateful to Lisa Weyant for her help in typesetting and formatting this document.

REFERENCES

Andrews, R.L., Srinivasan, T.C., 1995. Studying consideration effects in empirical choice models using scanner panel data. Journal of Marketing Research 32 (February), 30-41.

Ashford, N., Benchemam, M., 1987. Passengers' choice of airport: an application of the multinomial logit model. Transportation Research Record 1147, 1-5.

Ben-Akiva, M., Boccara, B., 1995. Discrete choice models with latent choice sets. International Journal of Research in Marketing 12, 9-24.

Börsch-Supan, A., 1987. Econometric analysis of discrete choice. Lecture Notes in Economics and Mathematical Systems, Springer-Verlag, Berlin, Germany, 34.

Bureau of Transportation Statistics, 1997. Airline On-time Flight Statistics.

Chiang, J., Chib, S., Narasimhan, C., 1999. Markov chain Monte Carlo and models of consideration set and parameter heterogeneity. Journal of Econometrics 89, 223-248.

Furuichi, M., Koppelman, F.S., 1994. An analysis of air travelers' departure airport and destination choice behavior. Transportation Research A 28 (3), 187-195.

Gensch, D., 1987. A two-stage disaggregate attribute choice model. Marketing Science 6 (3), 223-231.

Harvey, G., 1987. Airport choice in a multiple airport region. Transportation Research A 21 (6), 439-449.

Innes, J.D., Doucet, D.H., 1990. Effects of access distance and level of service on airport choice. Journal of Transportation Engineering 116 (4), 507-516.

Manski, C., 1977. The structure of random utility models. Theory and Decision 8, 229-254.

Manski, C., Lerman, S., 1977. The estimation of choice probabilities from choice-based samples. Econometrica 45 (8), 1977-1988.

Metropolitan Transportation Commission, 1995. Airline Passenger Survey. Oakland, CA.

Ndoh, N.N., Pitfield, D.E., Caves, R.E., 1990. Air transportation passenger route choice: a nested multinomial logit analysis. Spatial Choices and Processes. M.M. Fischer, P. Nijkamp, Y.Y. Papageorgiou (editors). Elsevier Science Publishers B.V., North Holland, 349-365.

Official Airline Guide, 1995. Market Analysis. OAG Worldwide, Oakbrook, IL.

Ozoka, A.I., Ashford, N., 1989. Application of disaggregate modeling in aviation systems planning in Nigeria: a case study. Transportation Research Record 1214, 10-20.

Pels, E., Nijkamp, P., Rietveld, P., 2001. Airport and airline choice in a multiple airport region: an empirical analysis for the San Francisco Bay Area. Regional Studies 35 (1), 1-9.

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

Roberts, J.H., Lattin, J.M., 1991. Development and testing of a model of consideration set composition. Journal of Marketing Research 28, 429-440.

Shocker, A.D., Ben-Akiva, M., Boccara, B., Nedungadi, P., 1991. Consideration set influences on consumer decision-making and choice: issues, models, and suggestions. Marketing Letters 2 (3), 181-197.

Skinner, R.E., Jr., 1976. Airport choice: an empirical study. Transportation Engineering Journal, 102 (4), 871-883.

Swait, J., 1984. Probabilistic choice set formation in transportation demand models. Unpublished Ph.D. Thesis, Department of Civil Engineering, MIT, Cambridge, MA.

Swait, J., 2001. Choice set generation within the generalized extreme value family of discrete choice models. Transportation Research B 35 (7), 643-666.

Swait, J., Ben-Akiva, M., 1987. Incorporating random constraints in discrete models of choice set generation. Transportation Research B 21 (2), 91-102.

Thompson, A., Caves, R., 1993. The projected market share for a new small airport in the north of England. Regional Studies 27 (2), 137-147.

Williams, H., Ortuzar, J., 1982. Behavioural theories of dispersion and the mis-specification of travel demand models. Transportation Research B 16 (3), 167-219.

Windle, R., Dresner, M., 1995. Airport choice in multiple-airport regions. Journal of Transportation Engineering 121 (4), 332-337.

LIST OF TABLES

Table 1. Estimation Sample Shares, Market Shares, and Weights

Airport	Estimation sample shares	Market shares	Weight[1]
San Francisco International (SFO)	0.2559	0.6248	2.4420
San Jose International (SJC)	0.4932	0.1775	0.3596
Oakland International (OAK)	0.2509	0.1977	0.7882

[1]The weight variable refers to the weight placed on individuals choosing each airport. Thus, for example, each individual in the estimation sample choosing SFO is assigned a weight of 2.4420 during estimation.

Table 2. Estimation Results

Variable	MNL Model		PCMNL Model			
			Consideration Stage		Choice Stage	
	Parameter	t-statistic	Parameter	t-statistic	Parameter	t-statistic
Access time-related variables (access time is in 100s of minutes)						
Access time	-6.964	-13.43	-2.185	-1.91	-7.503	-11.13
Access time x traveling alone	-0.825	-1.76	---	---	-2.169	-3.24
Access time x female	-0.796	-1.92	1.748	2.02	-0.701	-1.11
Access time x weekday travel	---	---	-3.788	-2.89	---	---
Frequency-related variables (frequency is in flights per day divided by 10)						
Frequency	0.411	2.88	3.893	5.83	0.360	2.19
Frequency x traveling alone	0.271	1.87	---	---	0.232	1.40
Frequency x female	-0.173	-1.36	---	---	-0.092	-0.62
Frequency x high income indicator (annual income > 150K)	-0.257	-2.09	---	---	-0.581	-2.85
Frequency x weekday travel	---	---	1.832	2.00	---	---
Airport Constants						
San Francisco International	---	---	3.826	3.08	---	---
San Jose International	-1.998	-12.30	-0.595	-1.90	-1.659	-8.44
Oakland International	-2.162	-17.17	-1.531	-3.14	-1.522	-10.98

Table 3. Time Value of Frequency of Service at Choice Stage

Population Subgroup	MNL	PCMNL[1, 2]
Male, low-income, traveling in a group	5.9	4.8
Male, high-income, traveling in a group	2.3	--
Male, low-income, traveling alone	8.8	6.1
Male, high-income, traveling alone	5.5	≈ 0
Female, low-income, traveling in a group	3.1	3.3
Female, high-income, traveling in a group	≈ 0	--
Female, low-income, traveling alone	5.9	4.8
Female, high-income, traveling alone	2.9	≈ 0

[1]The numbers indicate the additional access time travelers are willing to endure for an increase in ten flights per day to their destination.

[2]A "--" entry indicates that frequency has a negative effect at the choice stage for the corresponding population group. While not intuitive, these negative frequency effects are not significantly different from zero.

Table 4. Elasticity Effects of Quality of Service Improvements

Improvement in Quality of Service	Elasticity Impact on Market Share					
	MNL Model			PCMNL Model		
	SFO	SJC	OAK	SFO	SJC	OAK
San Francisco International (SFO)						
Decrease in travel time	1.313	-1.597	-2.715	0.870	-1.150	-1.709
Increase in air frequency	0.277	-0.393	-0.524	0.169	-0.237	-0.322
San Jose International (SJC)						
Decrease in travel time	-0.220	1.111	-0.301	-0.205	0.971	-0.223
Increase in air frequency	-0.054	0.227	-0.034	-0.096	0.400	-0.056
Oakland International (OAK)						
Decrease in travel time	-0.566	-0.306	2.063	-0.431	-0.251	1.582
Increase in air frequency	-0.114	-0.053	0.409	-0.152	-0.079	0.549

Table 5. Measures of Fit in Estimation and Validation Sample

Summary Statistic	Estimation Sample		Validation Sample	
	MNL	PCMNL	MNL	PCMNL
Log-likelihood at zero	-1777.55	-1777.55	-329.58	-329.58
Log-likelihood at market shares	-1490.40	-1490.40	-275.69	-275.69
Log-likelihood at convergence (estimation) / Predictive log-likelihood (validation)	-897.50	-697.04	-174.74	-151.15
Number of parameters[1]	7	15	7	15
Number of observations	1618	1618	300	300
Adjusted likelihood ratio index (estimation) / Predictive adjusted ratio index (validation)	0.393	0.522	0.340	0.397
Average probability of correct prediction	0.662	0.749	0.665	0.729

[1]The number of parameters refers to the coefficients on the exogenous variables; it does not include the alternative-specific constants in the MNL model and the alternative-specific constants at the choice stage of the PCMNL model.