Modeling the Competition among Air Travel Itinerary Shares: GEV Model Development

Gregory M. Coldren, Northwestern University & United Airlines
Frank S. Koppelman, Northwestern University

Conference paper
Session III

Moving through nets:
The physical and social dimensions of travel
10th International Conference on Travel Behaviour Research

Lucerne, 10-15. August 2003

MODELING THE COMPETITION AMONG AIR TRAVEL ITINERARY SHARES: GEV MODEL DEVELOPMENT

Gregory M. Coldren *(corresponding author)*
Northwestern University & United Airlines
Department of Civil Engineering
Robert R. McCormick School of Engineering and Applied Science
2145 Sheridan Road
Evanston, IL 60208-3109
Ph: (847) 491-7612
Fax: (847) 491-4011
Email: gregorycoldren@yahoo.com

Frank S. Koppelman, Ph.D.
Department of Civil Engineering
Robert R. McCormick School of Engineering and Applied Science
Northwestern University
2145 Sheridan Road
Evanston, IL 60208-3109
Ph: (847) 491-8794
Fax: (847) 491-4011
Email: f-koppelman@northwestern.edu

ABSTRACT

This study reports the results of aggregate air-travel itinerary share models estimated at the city-pair level for the top 500 East-West markets in the United States and Canada. These models predict airline ridership at the itinerary level and aid carriers in long and intermediate term decision-making. The models use official and comprehensive schedule and bookings data. GEV models are estimated capturing the itinerary share competition dynamic along three dimensions: time of day, carrier, and level-of-service. Models incorporate one, two, or three of these dimensions simultaneously. Model specifications include multinomial logit, one and two-level nested logit, and one and two-level weighted nested logit models. Independent variables for the models measure various itinerary service characteristics: level-of-service, connection quality, carrier, fare, aircraft type, and time of day. The results are intuitive, and the advanced models outperform the more basic specifications with regard to statistical tests and behavioral interpretations, giving insight into the competitive dynamic of air-carrier itineraries.

Keywords

Air Travel Demand Forecasting, Travel Behavior, GEV Models, Logit Share Modeling, International Conference on Travel Behaviour Research, IATBR

Preferred citation

Coldren, G. M., and F. S. Koppelman, (2003), Modeling the Competition among Air Travel Itinerary Shares: GEV Model Development, paper presented at the 10th International Conference on Travel Behaviour Research, Lucerne, August 2003.

1. INTRODUCTION

This paper reports the results of aggregate air-travel itinerary share models estimated at the city-pair level. These models forecast the number of passengers expected to travel on each itinerary between any city-pair conditional on the forecasted city-pair passenger volume. Thus, these models support air-carrier long and intermediate term decision-making with respect to fleet planning, merger and acquisition scenarios, route planning, code share opportunity scenarios, minimum connection time studies, price-elasticity studies, hub location studies etc. Itinerary share models provide carriers with an understanding of the relative importance of different service factors on city-pair itinerary share, and hence how policy changes to various itinerary service characteristics can increase itinerary (and therefore carrier) market share. Improvements to a carrier's itinerary share models leads to an improvement in its forecasting ability, which translates to improvements in revenue management, schedule efficiency, and profitability.

Value, the aggregate analog of utility, is used to represent the relative desirability of each itinerary connecting a city-pair. The market share assigned to each itinerary is modeled as a function of the value of the itinerary and the values of all other itineraries serving the city-pair for a given day of the week. Earlier work (Coldren et al., 2003) modeled itinerary shares as an aggregate multinomial logit (MNL) function of the itineraries' attributes. That work focused on the influence of different itinerary service characteristics on itinerary share (level-of-service, carrier, fare level, time of day, aircraft type etc.) However, due to the Independence of Irrelevant Alternatives (IIA) property of the MNL model (Koppelman et al., 2003; Ben-Akiva and Lerman, 1985), these models are believed to provide an incomplete representation of the substitution pattern among itineraries. The IIA property is unrealistic in the context of air-travel itinerary share modeling since it is likely that the competition among itineraries (as measured by cross-elasticities) is differentiated by proximity in departure time, level-of-service, carrier or any combination of these dimensions.

More advanced itinerary share models take account of the underlying competitive structure among city-pair itineraries. In particular, these models take account of the differential impact of changes in one itinerary on each other itinerary based on similarity of major itinerary characteristics such as time of day, carrier and/or level-of-service (nonstop, direct, single-connect and double-connect). Thus, the current study investigates the competition among itineraries within city-pairs that is not possible with the MNL model. We account for different substitution patterns between alternatives that are differentiated by proximity along the departure time, level-of-service, and/or carrier dimensions by developing several Generalized Extreme Value (GEV) models, which allow for the possibility of correlation between error terms for pairs of alternatives (McFadden, 1978; Ben-Akiva and Lerman, 1985; Koppelman and Sethi, 2000). These models estimate differential itinerary competition measurements within one model, capturing the competitive structure of itineraries simultaneously with the value function parameters.

We begin this paper with a "base" MNL model, which was developed through extensive estimation and validation testing and demonstrated to provide substantially improved itinerary share prediction than models previously used by a major air-carrier (Coldren et al., 2003). We then relax the constraints of the MNL model by use of one and two-level Nested Logit (NL) models (McFadden, 1978; Börsch-Supan, 1990). These models consider one, or a combination of two of the above proximity measures. We explore the structure and interpretation of these models, with particular emphasis on the analysis of the logsum parameters representing the level of itinerary competition within nests. These models are shown to outperform the base MNL model with respect to statistical tests and behavioral interpretations, leading to a clearer understanding of the air-travel itinerary competition dynamic.

Next, we propose and estimate one and two-level Weighted Nested Logit (WNL) models that combine the best results of the NL models. These models allow parallel consideration of more than one nesting structure with a weight indicating the relative importance of each nesting structure. The formulation of these models is similar to the formulation of Principles of Differentiation Models (Bresnahan et al., 1997; Koppelman and Sethi, 2000).

2. MODELING FRAMEWORK

Most studies of intercity air-travel have focused on air-carrier share at the system (Morash and Ozment, 1996; Suzuki et al., 2001), city-pair (Nason, 1981; Yoo and Ashford, 1996; Ghobrial and Soliman, 1992; Proussaloglou and Koppelman, 1995) or point-to-point (nonstop) flight share level (Algers and Beser, 1997; Proussaloglou and Koppelman, 1999). Few studies have employed itinerary-level data and/or focused on itinerary choice. The study reported here is based on comprehensive bookings and schedule data obtained from official sources that are linked to support estimation of itinerary shares for any city-pair. Bookings data was obtained from the Computer Reservation System (CRS), a data source containing detailed records of booked itineraries. CRS data contains the vast majority of commercial aviation bookings. However, increasing use of direct carrier booking has reduced the proportion of bookings reported by this source, a data problem that will have to be addressed in the future. Leg-based air-carrier schedule information was obtained from the Official Airline Guide (OAG Worldwide Limited, 2001). Finally, fare data was obtained from the "Superset" data source (Data Base Products, Inc., 2001). A more detailed explanation of these data sources (origins, available variables etc.) can be found in our earlier work estimating MNL itinerary share models (Coldren et al., 2003).

The motivation in developing our itinerary share models is to understand the underlying competitive structure of air-travel itineraries, as well as to determine the importance of various service attributes on itinerary choice. In this study, an itinerary is defined as a leg (flight number) or sequence of legs connecting a given city-pair. Itineraries are either nonstop, direct (a connecting itinerary involving no airplane change), single-connect (a connecting itinerary with an airplane change), or double-connect (an itinerary involving two connections). Itinerary choice is the fundamental air-traveler choice. Embedded in itinerary choice is the simultaneous choice of carrier, route, time of day and equipment. Once itinerary shares are predicted for all city-pairs, these forecasts can be "rolled-up" to predict carrier share at the flight-leg, city-pair, region, system, or any other level of aggregation.

United Airlines itinerary building engine was used to generate the set of feasible itineraries between the city-pairs using the leg-based schedule data. Itineraries are generated for each day of the week keeping in mind the days of the week that each itinerary's flight leg(s) operates. The dependent variables in the models are the number of passengers who booked each itinerary. This was determined by merging the generated itineraries with the CRS booked itineraries. Using May 2001 data, all models were estimated using maximum likelihood techniques for the top 500 city-pairs (as determined by inter-city air-travel volume) between the East and West (as determined by time zone) regions of the United States and Canada using the GAUSS modeling software (Aptech Systems, Inc., 2003). The choice (alternative) sets were modeled as the set of all itineraries between each city-pair for each day of the week. Please see our previous work for a more detailed description of the model data setup (Coldren et al., 2003).

Table 1 reports passenger and itinerary summary statistics of the top 100, 250, 500, 1000 and all (2481) East-West city-pairs. Based on the substantial reduction in computation time (proportional to the number of itineraries), and previous indications of good representation of results using subsets of the data, the top 500 markets were selected for estimation since they include well over 80% of the bookings. Tables 2-4 report the distribution of itineraries and booked passengers by level-of-service (Table 2), carrier (Table 3) and time of day (Table 4) for the estimation dataset. These data represent one week of service and travel across all 500 city-pair markets.

Variables that describe each itinerary, and their corresponding parameter estimates, determine an itinerary's value. Value is formulated as a weighted linear function of the explanatory variables. The variables used in our models are service characteristics describing each itinerary such as level-of-service indicators, connection quality measurements, fares, carrier constants, a code share indicator, aircraft type, and time of day as described in Table 5. Parameter estimates indicate the relative importance of different service factors on itinerary choice.

3. MULTINOMIAL LOGIT MODEL

The base or reference model for the study is the MNL model reported in Table 6. The parameter estimates are reported in groups corresponding to level-of-service (relative to the best level-of-service in the city-pair), connection quality, carrier attributes, aircraft type and time of day variables. These parameter estimates are very similar to the MNL parameter estimates from our previous study. That study details the interpretation of these value function parameter estimates and interested readers are referred to it (Coldren et al., 2003). For the current study, it is sufficient to state that all of the value function parameter estimates have the correct sign, are of reasonable magnitude and are statistically significant. Additionally, the MNL model explains a large portion (more than 27%) of the variability in the data.

The limitation of this model is that its IIA property implies that all itineraries within a city-pair-day-of-the-week compete equally with each other. However, our *a priori* belief is that itineraries sharing a common carrier, time of day and/or level-of-service will exhibit a greater degree of competition/substitution among themselves than with itineraries not sharing these attributes. We examine this expectation by estimating a sequence of one and two-level NL models and one and two-level WNL models with the same value function specification as the MNL model. Each of these models contains nests along the dimensions of time of day, level-of-service, and/or carrier. Within each dimension, please note that each itinerary belongs to one and only one nest. Finally, as can be seen from Tables 6-8, the value function parameter estimates are very similar across the different model specifications.

4. ONE-LEVEL NESTED LOGIT MODELS

Initial nested logit estimations assume one level of nesting based on one of the three dimensions described above. In these models, itineraries are grouped into nests according to time of day (morning, 5:00 - 9:59 A.M., midday, 10:00 A.M. - 3:59 P.M. and evening, 4:00 P.M. -

Midnight), carrier (six major carriers and a group of "other" carriers) or level-of-service. A visual representation of the one-level time NL model and the one-level carrier NL model is shown in Figures 1 and 2 respectively.

With a one-level nested logit specification, the share of passengers assigned to each itinerary between a city-pair for a given day of the week is given by:

$$S_i = S_{n'} \times S_{i|n'} = \frac{\exp\left(\frac{1}{\mu}\Gamma_n\right)}{\sum_{n' \in N} \exp\left(\frac{1}{\mu}\Gamma_{n'}\right)} \times \frac{\exp(\mu V_i)}{\sum_{i' \in n} \exp(\mu V_{i'})} \tag{1}$$

where S_i is the passenger share assigned to itinerary i,

$S_{n'}$ is the passenger share assigned to nest n',

$S_{i|n'}$ is the passenger share assigned to itinerary i given nest n',

μ is the logsum parameter associated with the nests,

$\Gamma_j = \ln\left(\sum_{i' \in N_j} \exp(\mu V_{i'})\right)$ and

V_i is the value of itinerary i.

The variance of the error for all alternatives is set equal to $\frac{\pi^2}{6}$ as is commonly done in MNL and NL models. For nested alternatives, the error associated with each alternative is decomposed into two components; a random component, independent for each alternative and a random component common to all alternatives in a nest. The error variance for each distinct alternative is given by $\frac{\pi^2}{6\mu^2}$, where μ is the logsum parameter. Thus, the logsum parameter must be greater than one to ensure that the component variance is smaller than the total variance. Larger values of the logsum parameter indicate a higher level of substitution within nests. Table 6 shows the estimation results for two one-level NL models nested by time of day and carrier[1]. These models reject the hypotheses that the MNL model is the true model. The one-level NL level-of-service model is not reported, as the logsum parameter was estimated to be less than one, which is inconsistent with utility theory.

Clearly, itineraries within a common time period and itineraries flown by the same carrier have common attributes (both included and excluded) that passengers consider in their itinerary selection process. On the other hand, grouping itineraries by level-of-service did not yield theoretically acceptable results. This was surprising since it seems likely that an itinerary within a given level-of-service nest (nonstop, direct, single-connect, or double-connect) would share many characteristics with the other alternatives within the same level-of-service nest and thus should have higher cross-elasticities among themselves than with itineraries of different levels-of-service. We will return to this issue in the following section.

5. TWO-LEVEL NESTED LOGIT MODELS

We estimated six two-level nested logit model specifications representing all possible two-level combinations for the three itinerary dimensions under study (upper-level time and lower-level carrier (time, carrier); carrier, time; time, level-of-service; level-of-service, time; carrier, level-of-service; and level-of-service, carrier). With a two-level nested logit specification, the share of passengers assigned to each itinerary between a city-pair for a given day of the week is given by the following equation:

$$
\begin{aligned}
S_i &= S_{m'} \times S_{n'|m'} \times S_{i|n'} \\
&= \frac{\exp\left(\frac{1}{\mu_M}\Gamma_m\right)}{\sum_{m'\in M}\exp\left(\frac{1}{\mu_M}\Gamma_{m'}\right)} \times \frac{\exp\left(\frac{\mu_M}{\mu_N}\Gamma_n\right)}{\sum_{n'\in N}\exp\left(\frac{\mu_M}{\mu_N}\Gamma_{n'}\right)} \times \frac{\exp(\mu V_i)}{\sum_{i'\in n}\exp(\mu V_{i'})}
\end{aligned}
\qquad (2)
$$

---

[1] Early experimental estimations yielded similar parameter estimates for the logsum variables across nests in each nesting dimension. Given this result and a desire for consistency across nests of the same type, we constrained the logsum parameter in each case to be equal across all common nests.

where $S_{m'}$ is the passenger share assigned to upper level nest m',

$S_{n'|m'}$ is the passenger share assigned to lower level nest n' given upper level nest m',

μ_M is the logsum parameter associated with the upper level nests,

μ_N is the logsum parameter associated with the lower level nests,

$$\Gamma_n = \ln\left(\sum_{i' \in N_j} \exp\left(\mu_N V_{i'} \right) \right) \text{ and}$$

$$\Gamma_m = \ln\left(\sum_{n' \in N_m} \exp\left(\frac{\mu_M}{\mu_N} \Gamma_{n'} \right) \right).$$

The total error variance for a two-level NL model can be decomposed into three components. The total error variance, $\frac{\pi^2}{6}$, is decomposed into a random component distinct to each alternative, $\frac{\pi^2}{6\mu_N{}^2}$; a random component distinct to each lower level nest but not including the error component of the elemental alternatives, $\frac{\pi^2}{6\mu_M{}^2} - \frac{\pi^2}{6\mu_N{}^2}$; and the balance associated with each upper level nest but excluding the error component of the lower level nest and the elemental alternative, $\frac{\pi^2}{6} - \frac{\pi^2}{6\mu_M{}^2}$. Thus, both of the logsum parameters must be greater than one, and the lower-level logsum parameter must be greater than the upper-level logsum parameter. Larger values of the logsum parameters indicate a higher rate of substitution among itineraries within a nest. The requirement that the lower-level logsum parameter be greater than the upper-level logsum parameter implies that itineraries within the same lower-level nest (and hence within the same upper-level nest) share the most attributes and compete more closely with each other than with other itineraries. Itineraries sharing a common upper-level nest (but not a lower-level nest) have a lower level of competition among themselves than itineraries within a lower-level nest, but a greater level of competition than with itineraries not contained in the upper-level nest. Finally, the level of competition among itineraries that do not share a common upper-level nest is the same as that found in the MNL model.

Of the six two-level models estimated, only two satisfied these logsum constraints. These models, reported in Table 7, are for time, level-of-service and time, carrier. The time, carrier model significantly rejects both the time and carrier one-level NL models. In addition, the time, level-of-service model significantly rejects the one-level time NL model. These results indicate that itineraries flown by the same carrier or of the same level-of-service, and within a limited time period, compete "strongly" with each other. While earlier one-level NL results indicated that itineraries flown by the same carrier were significantly nested, itineraries of the same level-of-service were not significantly nested across the entire day. That is, nonstop morning itineraries in a given city-pair exhibit a higher degree of substitution between themselves than between other nonstop itineraries in other time periods. This demonstrates the importance of conditioning the level-of-service competition dynamic by time period. A visual representation of the two-level time, level-of-service NL model is shown in Figure 3 and the two-level time, carrier NL model is shown in Figure 4.

The results indicate that there is moderate itinerary competition among itineraries sharing a common time period and greater competition among alternatives sharing both time period and carrier or time period and level-of-service. In the following sections, we propose and test models that combine the best one and two-level NL models to obtain the substitution relationships included in each of these pairs of models.

6. ONE-LEVEL WEIGHTED NESTED LOGIT MODEL

The one-level Weighted Nested Logit model simultaneously estimates a model with two parallel nest structures with a weighting parameter to indicate the relative importance of each nest. Each itinerary in each choice set appears twice in the model; once in each of the parallel nests. In this model, each nest is equivalent to a one-level NL model.

Due to the strong empirical results from the one-level time NL and the one-level carrier NL models, we estimated a one-level WNL model with a time nest and a carrier nest (see Figure 5). This model can be shown to be a special case of the Generalized Nested Logit (GNL) model

(Wen and Koppelman, 2000) or can be derived directly from the GEV construct using the generation function

$$G = p_t \sum_{t=1}^{T} \left(\sum_{j=1}^{J_t} Y_{jt}^{\mu_t} \right)^{\frac{1}{\mu_t}} + p_c \sum_{c=1}^{C} \left(\sum_{j=1}^{J_c} Y_{jc}^{\mu_c} \right)^{\frac{1}{\mu_c}} \tag{3}$$

The share of passengers assigned to each itinerary between a city-pair for a given day of the week is

$$\begin{aligned} S_i &= p_t \times S_t \times S_{i|t} + p_c \times S_c \times S_{ic} \\ &= p_t \times \frac{\exp\left(\frac{1}{\mu}\Gamma_t\right)}{\sum_{t' \in N} \exp\left(\frac{1}{\mu}\Gamma_{t'}\right)} \times \frac{\exp(\mu V_i)}{\sum_{i' \in t} \exp(\mu V_{i'})} \\ &+ p_c \times \frac{\exp\left(\frac{1}{\mu}\Gamma_c\right)}{\sum_{c' \in N} \exp\left(\frac{1}{\mu}\Gamma_{c'}\right)} \times \frac{\exp(\mu V_i)}{\sum_{i' \in c} \exp(\mu V_{i'})} \end{aligned} \tag{4}$$

where c represents the carrier nest,

t represents the time nest,

p_c is the weight given to the carrier nest and

$p_t = 1 - p_c$ is the weight given to the carrier nest

Estimation results for this model are reported in Table 8. The logsum parameters for both the time and carrier nests are significantly different than one, indicating increased itinerary competition among itineraries sharing a common time period or carrier. The weight parameter is close to ½ and significantly different than zero or one indicating that each portion of the structure is important. Finally, as can be seen from the log-likelihood values, this model outperforms both the one-level time and one-level carrier NL models, but it did not do as well as the two-level time, carrier NL model. This suggests that the substitution between itineraries flown by a carrier within a time period is more important than between carriers not confined to a time period.

7. TWO-LEVEL WEIGHTED NESTED LOGIT MODEL

The two-level weighted nested logit model is a direct extension of the one-level weighted nested logit model. The mathematical structure of the model is:

$$\begin{aligned}
S_i &= p_{ct} \times S_{ict} + p_{st} \times S_{ist} \\
&= p_{ct} \times S_{t_1} \times S_{c|t_1} \times S_{i|ct_1} + p_{st} \times S_{t_2} \times S_{s|t_2} \times S_{i|st_2} \\
&= p_{ct} \times \frac{\exp\left(\frac{1}{\mu_{T_1}}\Gamma_t\right)}{\sum_{t'\in T}\exp\left(\frac{1}{\mu_{T_1}}\Gamma_{t'}\right)} \times \frac{\exp\left(\frac{\mu_{T_1}}{\mu_C}\Gamma_c\right)}{\sum_{c'\in C}\exp\left(\frac{\mu_{T_1}}{\mu_C}\Gamma_{c'}\right)} \times \frac{\exp(\mu_C V_i)}{\sum_{i'\in c}\exp(\mu_C V_{i'})} \\
&\quad + p_{st} \times \frac{\exp\left(\frac{1}{\mu_{T_2}}\Gamma_t\right)}{\sum_{t'\in T}\exp\left(\frac{1}{\mu_{T_2}}\Gamma_{t'}\right)} \times \frac{\exp\left(\frac{\mu_{T_2}}{\mu_S}\Gamma_s\right)}{\sum_{s'\in S}\exp\left(\frac{\mu_{T_2}}{\mu_S}\Gamma_{s'}\right)} \times \frac{\exp(\mu_S V_i)}{\sum_{i'\in s}\exp(\mu_S V_{i'})}
\end{aligned} \tag{5}$$

Due to the significance of the two-level time, level-of-service and the two-level time, carrier NL models, we estimated a two-level WNL model with parallel two-level nests for time, carrier and time, level-of-service. Figure 6 gives a visual representation of this model and the estimation results are reported in Table 8. The logsum parameters for both time and level-of-service in the time, level-of-service nest are significantly larger than one and the level-of-service nest logsum is significantly larger than the time logsum parameter indicating a strong competitive structure within time periods and for itineraries with common level-of-service within time periods. The time logsum parameter in the time, carrier nest is almost identically equal to one (which does not support nesting) but the logsum parameter for carriers is significant indicating a high level of competition among itineraries flown by a carrier within a time period. The model is marginally better than the two-level time, carrier NL and the weight on the time, carrier nest is significantly different from one indicating that both nests are important. Further, this model is the only one that takes account of all three dimensions (time, level-of-service and carrier). Even though the value of the logsum for time in the time, carrier nest is not significant, the improvement in the model demonstrates the potential to obtain more interesting and significant improvements in model goodness of fit and structure through more complex nesting, implying that further exploration of the Weighted Nested Logit model is desirable.

8. STATISTICAL "DEFLATION" DUE TO DATA AGGREGATION

The statistical analysis is based on the choices of individual travelers; however, no individual data (socioeconomic, demographic etc.) is available to identify differences among travelers between any city-pair. Thus, it is questionable if the full weight of all these individual observations should be used in calculating the statistics for these models. The most extreme adjustment would be to count one for each alternative set and weight the alternative sets by the number of passengers in each. This can be accomplished after the fact by simply dividing the log-likelihood values by the ratio of the number of booked passengers to the number of city-pair-day-of-the-week combinations (380,593 / 3,493 = 108.96) and the t-statistics by the square root of that ratio (sqrt(108.96) = 10.44).

The statistical results that become non-significant under this adjustment are the rejection of the one-level NL time model by the two-level time, level-of-service NL model; and the rejection of the two-level time, carrier NL model by the two-level WNL model (the upper-level time logsum parameter estimate on the time, level-of-service side of the two-level WNL model also becomes insignificant). However, it is important to recognize that this adjustment is the most extreme possible adjustment. Intermediate adjustments that take account that we observe actual individual choice behavior would result in maintaining significance between these pairs of models.

9. CONCLUSIONS

This paper proposes and demonstrates the use of a further variation in the GEV family of models, the weighted nested logit (WNL) model, and shows that it has advantages over the somewhat more restrictive NL model structure. More generally, these results indicate that there is substantial room for additional model development within the GEV structure. Such development offers the opportunity to address a wide variety of competitive relationships in any choice set while retaining the advantages of a closed form model.

10. REFERENCES

Algers, S., and M. Beser (1997) A Model for Air Passengers Choice of Flight and Booking Class – a Combined Stated Preference and Reveled Preference Approach, ATRG Conference Proceedings. Vancouver, Canada.

Ben-Akiva, M. E., and S. R. Lerman (1985) *Discrete Choice Analysis: Theory and Application to Travel Demand*. The MIT Press. Cambridge, Massachusetts.

Börsch-Supan, A. (1990) On the Compatibility of Nested Logit Models with Utility Maximization. *Journal of Econometrics*, Vol.**32**, pp.371-387.

Bresnahan, T. F., S. Stern, and M. Trajtenberg (1997) Market Segmentation and the Sources of Rents from Innovation: Personal Computers in the Late 1980's. *RAND Journal of Economics*, Vol. **28**, No. 0, pp. S17-S44.

Coldren, G. M., and F. S. Koppelman. Modeling Air Travel Itinerary Shares: Logit Model Development at a Major U.S. Airline. *Journal of Air Transport Management*, In Press.

Daly, A., and S. Zachary (1978) Improved Multiple Choice Models. In *Determinants of Travel Choice* (Edited by D. A. Hensher and M. Q. Dalvi), Saxon House, Sussex.

Forinash, C. V., and F. S. Koppelman (1993). Application and Interpretation of Nested Logit Models of Intercity Mode Choice. *Transportation Research Record*, No. **1413**, pp. 98-106.

Ghobrial, A., and S. Y. Soliman (1992) An Assessment of Some Factors Influencing the Competitive Strategies of Airlines in Domestic Markets. *International Journal of Transport Economics*, Vol. **19**, No. 3, pp. 247-258.

Koppelman, F. S., and V. Sethi (2000) Closed-Form Discrete-Choice Models, In *Handbook of Transport Modelling*, Hensher, D.A. and K. J. Button, eds., Elsevier Science Ltd., Amsterdam.

Koppelman, F. S., C. Bhat, V. Sethi, and B. Williams (2003) *A Self Instructing Course in Mode Choice Modeling*, U.S. Department of Transportation, Federal Highway and Federal Transit Administrations.

McFadden, D. (1978) Modeling the Choice of Residential Location. *Transportation Research Record*, No. **672**, pp. 72-77.

Morash, E. A., and J. Ozment (1996) The Strategic Use of Transportation Time and Reliability for Competitive Advantage. *Transportation Journal*, Vol. **36**, No. 2, pp. 35-46.

Nason, S. D. (1981) The Airline Preference Problem: An Application of Disaggregate Logit. Presented at the AGIFORS Symposium. Santa Barbara, California.

Proussaloglou, K.E., and F. S. Koppelman (1995) Air Carrier Demand: An Analysis of Market Share Determinants. *Transportation*, Vol. **22**, pp. 371-388.

Proussaloglou, K.E., and F. S. Koppelman (1999) The Choice of Carrier, Flight and Fare Class. *Journal of Air Transport Management*, Vol. **5**, No. 4, pp. 193-201.

Suzuki, Y., J. Tyworth, and R. Novack (2001) Airline Market Share and Customer Service Quality: a Reference-dependent Model. *Transportation Research A*, Vol. **35**, pp. 773-788.

Wen, C-H., and F. S. Koppelman (2001) The generalized nested logit model. *Transportation Research-B*, Vol. **35B**, No. 7, pp. 627-641.

Yoo, K., and N. Ashford (1996) Carrier Choices of Air Passengers in Pacific Rim: Using Comparative Analysis and Complementary Interpretation of Revealed Preference and Stated Preference Data. *Transportation Research Record*, No. **1562**, pp. 1-7.

OAG Worldwide Limited (2001) *Official Airline Guide*. Bedfordshire, LU5 4HB, United Kingdom.

Data Base Products, Inc. (2001) *Superset*. Dallas, TX, USA.

Aptech Systems, Inc. (2003) *GAUSS*. Maple Valley, WA, USA.

ACKNOWLEDGMENT

Laurie Garrow, a current Civil Engineering graduate student at Northwestern University, provided very useful help with coding the constraints used in the models. The authors are thankful for her support.

LIST OF TABLES AND FIGURES

TABLE 1. Passenger and Itinerary Summary Statistics for the East-West Region

Number of Largest East-West City-Pairs	Percentage of Booked Passengers for Region (2481 Markets)	Percentage of Itineraries for Region
100	49	18
250	70	35
500	84	52
1000	92	73
2481	100	100

TABLE 2. Number of Itineraries and Booked Passengers by Level-of-Service for the Top 500 East-West City-Pairs (all days of the week combined)

Level-of-Service	Itineraries	Booked Passengers
Nonstop	3,170	134,248
Direct	3,395	16,278
Single-Connect	178,289	228,004
Double-Connect	151,825	2,063
Total	**336,679**	**380,593**

TABLE 3. Number of Itineraries and Booked Passengers by Carrier for the Top 500 East-West City-Pairs (all days of the week combined)

Carrier	Itineraries	Booked Passengers
United	60,045	75,390
American	49,517	61,187
Continental	50,946	39,118
Delta	48,587	56,207
Northwest	41,303	32,242
U.S. Airways	20,762	33,521
Other	65,519	82,928
Total	**336,679**	**380,593**

TABLE 4. Number of Itineraries and Booked Passengers by Time of Day for the Top 500 East-West City-Pairs (all days of the week combined)

Time of Day	Itineraries	Booked Passengers
5-6 A.M.	2,866	1,888
6-7 A.M.	46,793	32,923
7-8 A.M.	34,206	43,115
8-9 A.M.	22,105	32,404
9-10 A.M.	26,619	30,150
10-11 A.M.	18,372	15,138
11-12 Noon	20,028	17,097
12-1 P.M.	25,484	22,192
1-2 P.M.	27,837	18,150
2-3 P.M.	24,149	17,335
3-4 P.M.	23,407	23,066
4-5 P.M.	24,182	26,231
5-6 P.M.	22,097	38,301
6-7 P.M.	10,102	26,839
7-8 P.M.	5,915	18,076
8-9 P.M.	2,164	10,670
9-10 P.M.	291	4,996
10-12 Midnight	62	2,022
Total	**336,679**	**380,593**

TABLE 5. Description of Explanatory Variables

Variable	**Description**
Level-of-Service	Dummy variable representing the level-of-service of the itinerary (nonstop, direct, single-connect, or double-connect) with respect to the best level-of-service available in the city-pair.
Second-Best Connection	For connection itineraries sharing a common leg, a dummy variable indicating that the itinerary is not the best connection (with respect to ground time) for the given incoming or outgoing leg at a transfer city.
Second-Best Connection Time Difference	If the second-best connection indicator equals one, this variable measures the ground time difference between the itinerary and the best connection itinerary.
Distance Ratio	Itinerary distance divided by the shortest itinerary distance for the city-pair multiplied by 100.
Fare Ratio	Carrier average fare divided by the industry average fare for the city-pair multiplied by 100.
Carrier	Dummy variable representing major U.S. domestic carriers. All other carriers are combined together in a single category.
Code share	Dummy variable indicating whether any leg of the itinerary was booked as a code share.
Regional Jet	Dummy variable indicating whether the smallest aircraft on any part of the itinerary is a regional jet.
Propeller Aircraft	Dummy variable indicating whether the smallest aircraft on any part of the itinerary is a propeller aircraft.
Time of Day	Dummy variable for each hour of the day (based on the local departure time of the first leg of the itinerary).

TABLE 6. Itinerary Share Models: MNL and One-Level NL's

Explanatory Variables	**Model**		
	MNL	**1-Level NL Time**	**1-Level NL Carrier**
Level-of-Service			
Nonstop Itinerary in Nonstop Market	0.0000	0.0000	0.0000
Direct Itinerary in Nonstop Market	-2.1665	-1.8137	-1.9705
Single-Connect Itinerary in Nonstop Market	-3.0202	-2.5282	-2.7413
Double-Connect Itinerary in Nonstop Market	-6.8665	-5.7235	-6.0281
Direct Itinerary in Direct Market	0.0000	0.0000	0.0000
Single-Connect Itinerary in Direct Market	-0.8587	-0.7241	-0.7677
Double-Connect Itinerary in Direct Market	-4.5000	-3.7316	-3.9348
Single-Connect Itinerary in Single-Connect Market	0.0000	0.0000	0.0000
Double-Connect Itinerary in Single-Connect Market	-3.4377	-2.8477	-3.0116
Connection Quality			
Second-Best Connection	-0.1418	-0.1020	-0.1058
Second-Best Connection Time Difference	-0.0126	-0.0104	-0.0107
Distance Ratio	-0.0299	-0.0247	-0.0279
Carrier Attributes			
Fare Ratio	-0.0069	-0.0059	-0.0045
Carrier Constants (Proprietary)	--------	--------	--------
Code share	-1.9828	-1.6334	-1.7680
Aircraft Type			
Mainline Jet	0.0000	0.0000	0.0000
Regional Jet	-0.5015	-0.4214	-0.4643
Propeller Aircraft	-0.5820	-0.4823	-0.5153
Time of Day			
5 - 6 A.M.	-0.2659	-0.2258	-0.2410
6 - 7 A.M.	0.0000	0.0000	0.0000
7 - 8 A.M.	0.0976	0.0757	0.0860
8 - 9 A.M.	0.2194	0.1725	0.1971
9 -10 A.M.	0.2395	0.1915	0.2125
10 - 11 A.M.	0.2543	0.1714	0.2288
11 - 12 noon	0.1900	0.1265	0.1813
12 - 1 P.M.	0.2316	0.1609	0.2144
1 - 2 P.M.	0.1081	0.0567	0.1045
2 - 3 P.M.	0.1085	0.0509	0.1046
3 - 4 P.M.	0.1915	0.1228	0.1745
4 - 5 P.M.	0.0684	0.0305	0.0679
5 - 6 P.M.	0.1227	0.0741	0.1034
6 - 7 P.M.	0.1940	0.1402	0.1650
7 - 8 P.M.	0.1210	0.0818	0.1186
8 - 9 P.M.	-0.0539	-0.0543	-0.0318
9 - 10 P.M.	-0.2932	-0.2448	-0.2455
10 – Midnight	-0.3181	-0.2981	-0.2686
Logsum Parameter	--------	1.2138	1.1699
Log Likelihood at Zero	-1,860,747	-1,860,747	-1,860,747
Log Likelihood at Convergence	-1,341,583	-1,341,093	-1,340,459
Rho-square w.r.t. Zero	0.2790	0.2793	0.2796

All variables are statistically significant at 95% confidence level

TABLE 7. Itinerary Share Models: Two-Level NL's

Explanatory Variables	Model	
	2-Level NL: Time, LOS	2-Level NL: Time, Carrier
Level-of-Service		
Nonstop Itinerary in Nonstop Market	0.0000	0.0000
Direct Itinerary in Nonstop Market	-1.8730	-1.8513
Single-Connect Itinerary in Nonstop Market	-2.4786	-2.5621
Double-Connect Itinerary in Nonstop Market	-5.8475	-5.5131
Direct Itinerary in Direct Market	0.0000	0.0000
Single-Connect Itinerary in Direct Market	-0.6356	-0.7397
Double-Connect Itinerary in Direct Market	-3.7683	-3.6209
Single-Connect Itinerary in Single-Connect Market	0.0000	0.0000
Double-Connect Itinerary in Single-Connect Market	-2.9497	-2.7763
Connection Quality		
Second-Best Connection	-0.0957	-0.0583
Second-Best Connection Time Difference	-0.0101	-0.0096
Distance Ratio	-0.0240	-0.0255
Carrier Attributes		
Fare Ratio	-0.0057	-0.0040
Carrier Constants (Proprietary)	--------	--------
Code share	-1.5797	-1.6116
Aircraft Type		
Mainline Jet	0.0000	0.0000
Regional Jet	-0.4105	-0.4387
Propeller Aircraft	-0.4656	-0.4648
Time of Day		
5 - 6 A.M.	-0.2209	-0.2450
6 - 7 A.M.	0.0000	0.0000
7 - 8 A.M.	0.0705	0.0784
8 - 9 A.M.	0.1639	0.1735
9 -10 A.M.	0.1833	0.1902
10 - 11 A.M.	0.1644	0.1745
11 - 12 noon	0.1214	0.1410
12 - 1 P.M.	0.1544	0.1666
1 - 2 P.M.	0.0537	0.0688
2 - 3 P.M.	0.0478	0.0587
3 - 4 P.M.	0.1163	0.1326
4 - 5 P.M.	0.0224	0.0380
5 - 6 P.M.	0.0635	0.0594
6 - 7 P.M.	0.1278	0.1227
7 - 8 P.M.	0.0731	0.0928
8 - 9 P.M.	-0.0560	-0.0282
9 - 10 P.M.	-0.2392	-0.2148
10 – Midnight	-0.2914	-0.2620
Upper-Level Logsum Parameter	1.1786	1.0587
Lower-Level Logsum Parameter	1.2546	1.3408
Log Likelihood at Zero	-1,860,747	-1,860,747
Log Likelihood at Convergence	-1,341,040	-1,338,513
Rho-square w.r.t. Zero	0.2793	0.2807

All variables are statistically significant at 95% confidence level

TABLE 8. Itinerary Share Models: One and Two-Level WNL's

Explanatory Variables	**Model** **1-Level WNL: Time | Carrier**	 **2-Level WNL: Time | Carrier , Time | LOS**
Level-of-Service		
Nonstop Itinerary in Nonstop Market	0.0000	0.0000
Direct Itinerary in Nonstop Market	-1.5670	-1.8681
Single-Connect Itinerary in Nonstop Market	-2.1916	-2.5381
Double-Connect Itinerary in Nonstop Market	-4.8582	-5.5631
Direct Itinerary in Direct Market	0.0000	0.0000
Single-Connect Itinerary in Direct Market	-0.6120	-0.6826
Double-Connect Itinerary in Direct Market	-3.1462	-3.6279
Single-Connect Itinerary in Single-Connect Market	0.0000	0.0000
Double-Connect Itinerary in Single-Connect Market	-2.4181	-2.8289
Connection Quality		
Second-Best Connection	-0.0632	-0.0528
Second-Best Connection Time Difference	-0.0086	-0.0095
Distance Ratio	-0.0223	-0.0252
Carrier Attributes		
Fare Ratio	-0.0037	-0.0040
Carrier Constants (Proprietary)	--------	--------
Code share	-1.3949	-1.5889
Aircraft Type		
Mainline Jet	0.0000	0.0000
Regional Jet	-0.3744	-0.4319
Propeller Aircraft	-0.4045	-0.4541
Time of Day		
5 - 6 A.M.	-0.1983	-0.2417
6 - 7 A.M.	0.0000	0.0000
7 - 8 A.M.	0.0630	0.0753
8 - 9 A.M.	0.1471	0.1679
9 -10 A.M.	0.1623	0.1859
10 - 11 A.M.	0.1436	0.1703
11 - 12 noon	0.1136	0.1373
12 - 1 P.M.	0.1428	0.1635
1 - 2 P.M.	0.0528	0.0668
2 - 3 P.M.	0.0461	0.0566
3 - 4 P.M.	0.1065	0.1293
4 - 5 P.M.	0.0360	0.0348
5 - 6 P.M.	0.0602	0.0546
6 - 7 P.M.	0.1144	0.1161
7 - 8 P.M.	0.0810	0.0883
8 - 9 P.M.	-0.0298	-0.0310
9 - 10 P.M.	-0.1919	-0.2107
10 – Midnight	-0.2353	-0.2564
Upper-Level Logsum Parameter (Time)	1.6055	--------
Upper-Level Logsum Parameter (Carrier)	1.3971	--------
Upper-Level Logsum Parameter (Time)	--------	1.0000
Lower-Level Logsum Parameter (Carrier)	--------	1.3249
Upper-Level Logsum Parameter (Time)	--------	1.3125
Lower-Level Logsum Parameter (LOS)	--------	1.5824
Weight Parameter	0.4944	0.8247
Log Likelihood at Zero	-1,860,747	-1,860,747
Log Likelihood at Convergence	-1,339,699	-1,338,477
Rho-square w.r.t. Zero	0.2800	0.2807

All variables are statistically significant at 95% confidence level

FIGURE 1. One-Level NL Time Model Structure

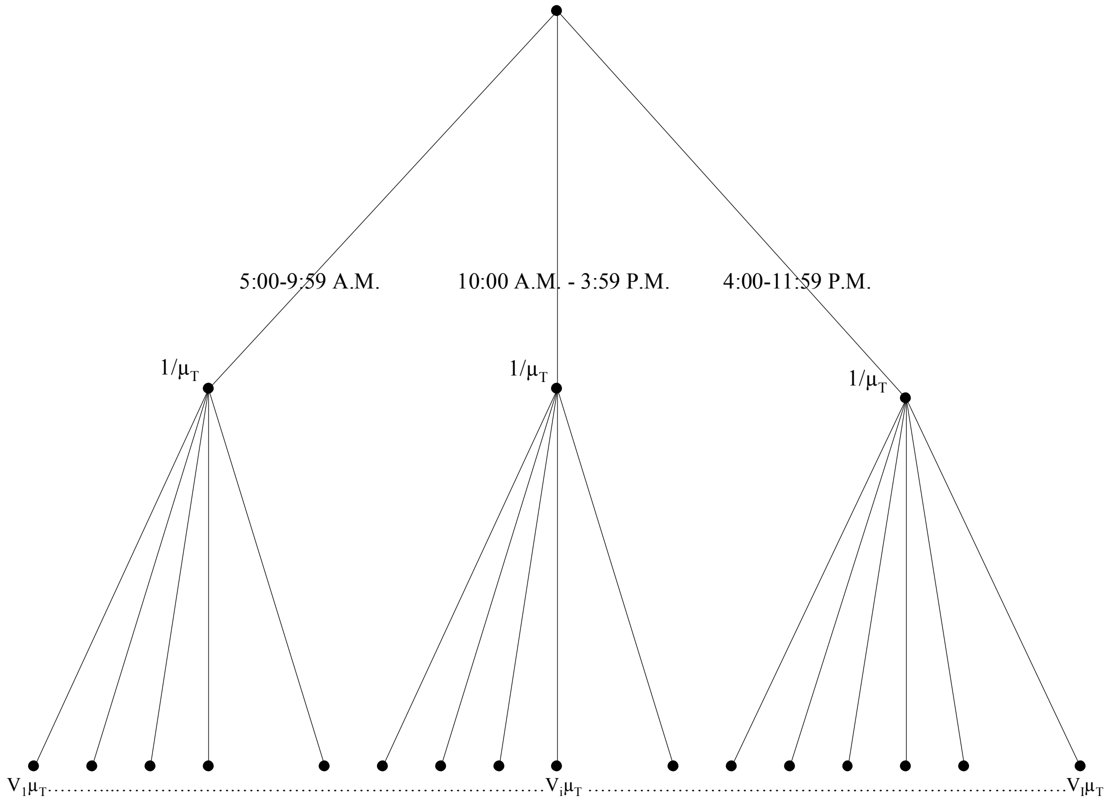

FIGURE 2. One-Level NL Carrier Model Structure

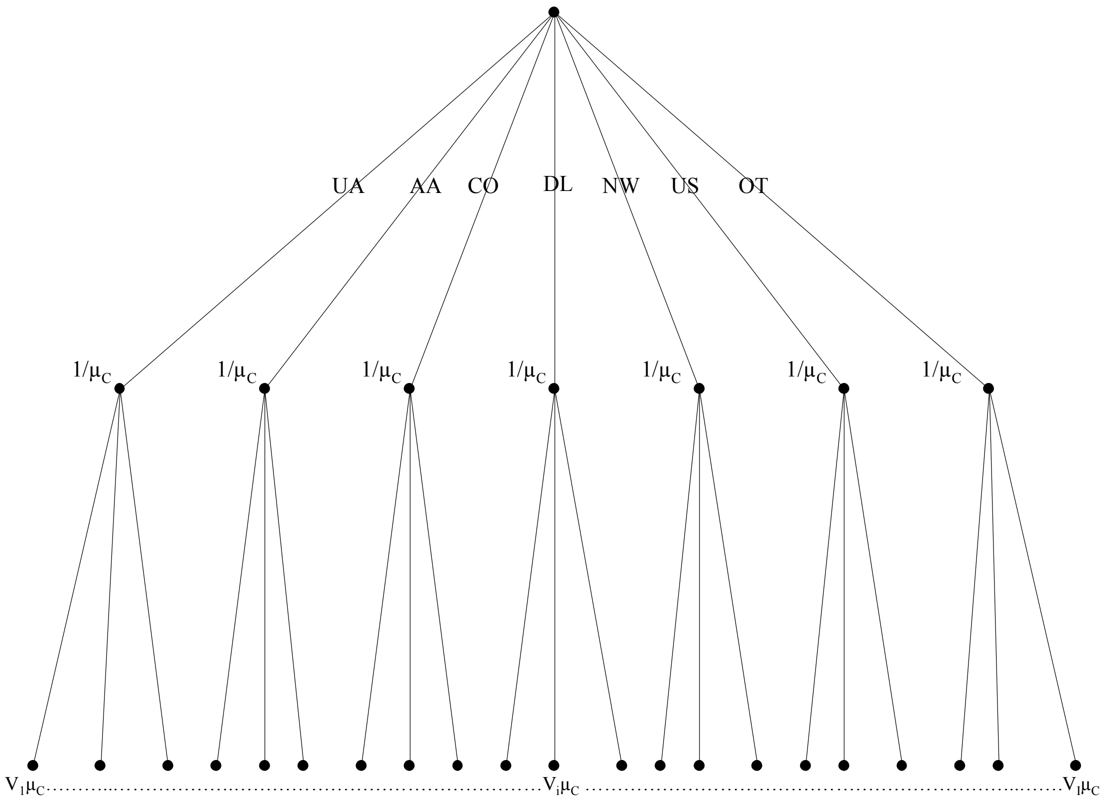

FIGURE 3. Two-Level NL Time, Level-of-Service Model Structure

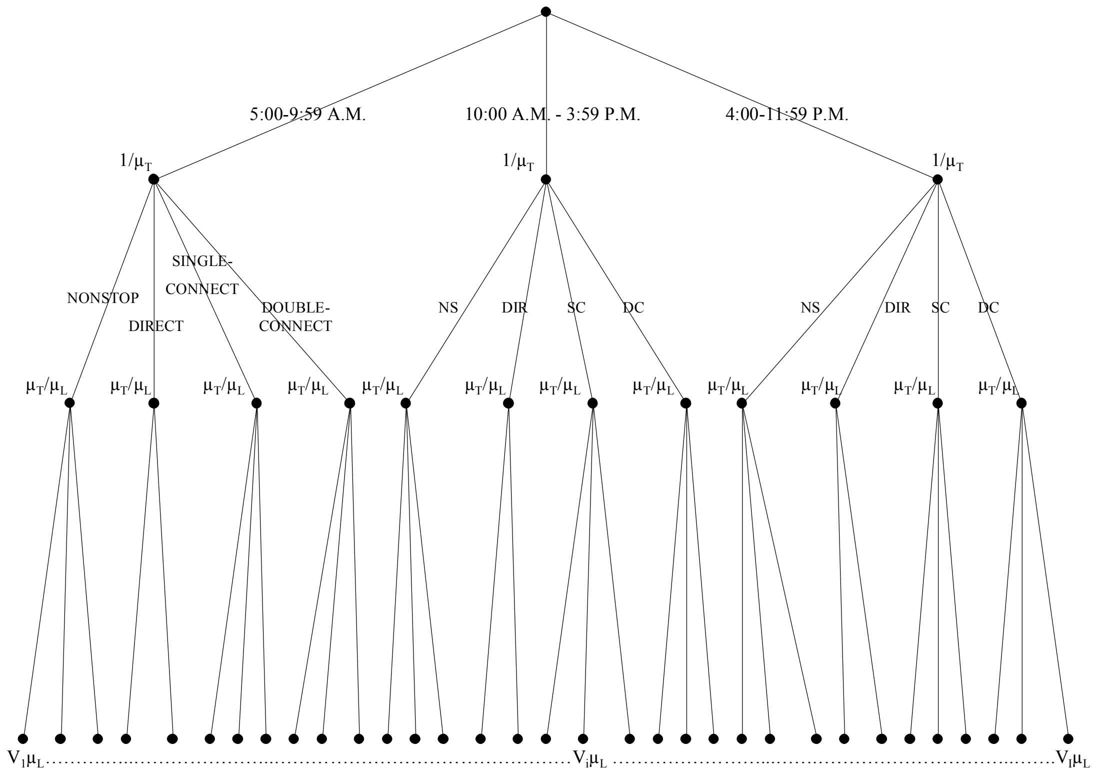

FIGURE 4. Two-Level NL Time, Carrier Model Structure

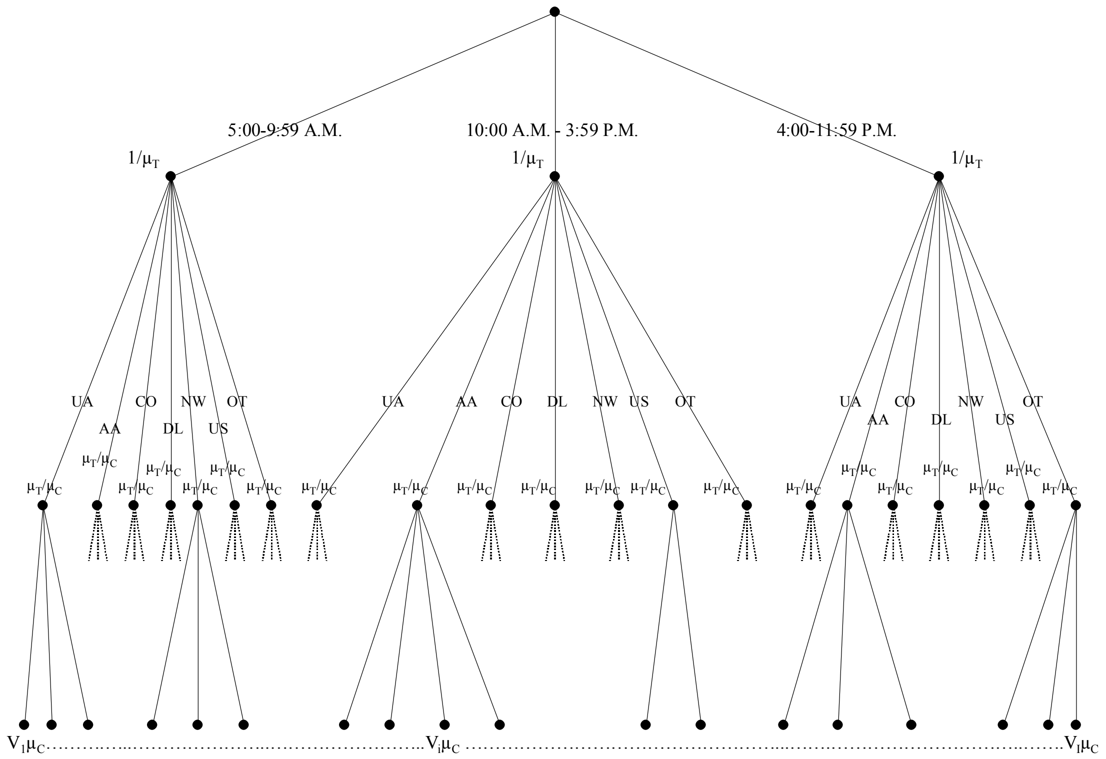

FIGURE 5. One-Level WNL Time | Carrier Model Structure

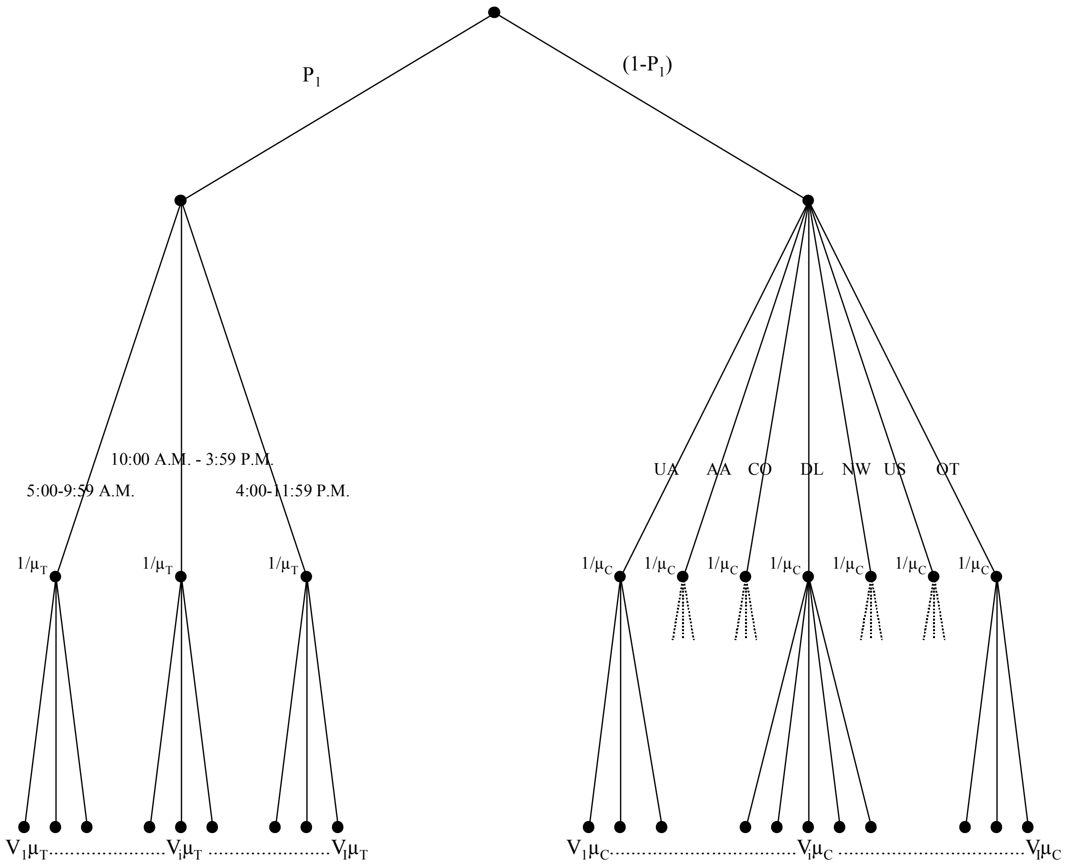

FIGURE 6. Two-Level WNL Time | Carrier, Time | Level-of-Service Model Structure