Journal of Air Transport Management 32 (2013) 8–16

Contents lists available at SciVerse ScienceDirect

Journal of Air Transport Management

journal homepage: www.elsevier.com/locate/jairtraman

Modelling air carrier choices with a Segment Specific Cross Nested Logit model

Tomasz Drabas*, Cheng-Lung Wu

School of Aviation, The University of New South Wales, Sydney, NSW 2052, Australia

Keywords:
Cross-nested choice model
Market segmentation
Brand loyalty
Latent class models

A B S T R A C T

A number of customer segments may exist in any air travel market, differing in the willingness-to-pay, income, age, time of travel or airline preferences. Not accounting for this diversity may introduce a significant variance into any choice model. In this paper we present a Segment Specific Cross-Nested Logit with Brand Loyalty (SSCNL-BL) model that explicitly accounts for this heterogeneity. The model is estimated using data from a stated preference choice experiment conducted among Australians travelling to the United States. The resulting SSCNL-BL model performs better than the more complex Mixed Logit model without the additional computational burden.

1. Introduction

The deregulation of the United States airline market in 1978 transformed the business substantially, both locally and globally. The implementation of more sophisticated methods of managing seats that followed has made the air travel more affordable. Today, thanks to tickets being sold for as low as $33[1] on the Sydney to Melbourne route, almost everyone can travel. In addition, consumers no longer rely on limited options presented by a booking agent; the Internet, at virtually no cost,[2] allows potential passengers to weigh up various travel options and choose the one they like the most. However, many factors influence the choice of a potential traveller: different spending power, travel purpose or past experience will render divergent attitudes towards price of the offer or the carrier.

These differences in choice makers' characteristics have a potential to introduce significant variance into any choice model, and render misleading or statistically insignificant parameter estimates. Segmenting a market is one of the ways of mitigating this problem. Greene and Hensher (2003) described an extension to the Multinomial Logit Model (MNL) based on the latent class formulation. Their model relaxed the requirement of making specific assumptions about the distribution of the parameters across the individuals. They found that their model outperformed the Mixed Logit (ML) model but argued that their findings were inconclusive and in a different settings ML could be more suitable.

A very similar approach to Greene and Hensher's development was investigated by Carrier (2008a). Carrier analysed airline itinerary and fare product choices of passengers travelling between Paris and three destinations in Germany. He used booking data to develop a latent class segmentation model and extended MNL to account for varying tastes of passengers. He also presented a new approach to time-of-day preferences treatment. This approach showed improved results over models that used neither segmentation nor the continuous time-of-day preference function.

However, the models researched by Greene and Hensher, and Carrier did not account for inter-product correlations as MNL is incapable of dealing with such. In an air market with multitude of flights the likelihood of violating the MNL model's Independence from Irrelevant Alternatives (IIA) property (McFadden, 1974) can be high as some of the air travel options can share common unobserved characteristics. This limitation of MNL can be alleviated by either moving to ML or Error Components (EC) models (Hess, 2007; Train, 2009), or by grouping similar travel options into nests in a Nested Logit (NL) framework (see Ben Akiva and Lerman (1985); Garrow (2010); McFadden (1978); Train (2009)). However, both approaches have their drawbacks: ML and EC are flexible but computationally intensive, while NL still exhibits the IIA property within nests.

Our model extends the Cross-Nested Logit (CNL) through the introduction of segment specific coefficients. Unlike the above models which accounted only for personal tastes heterogeneity,

* Corresponding author. Tel.: +61 410 386 456.
E-mail addresses: t.drabas@unsw.edu.au, drabas.t@gmail.com (T. Drabas), c.l.wu@unsw.edu.au (C.-L. Wu).

[1] Source: www.expedia.com, one-way, departure June 26, 2013, checked on April 6, 2013.

[2] We are neglecting the time investment or the cognitive burden associated with using the Internet as for many people these factors would be negligible.

0969-6997/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jairtraman.2013.04.004

our model accounts for the covariance between products and between market segments.

CNL appears to be closely related to the Ordered GEV model presented by Small (1987). The name Cross-Nested Logit is acknowledged to Vovsha (1997) who applied it to model travel mode choice in Israel. Wen and Koppelman (2001) proposed a Generalized Nested Logit (GNL) model of which CNL is a straightforward restriction with only the assumptions about the equality of the logsum parameters being different. Papola (2004) reformulated the CNL model deriving it from Generalized Extreme Value (GEV) model. Papola provided a general formulation of the model covariance matrix and a complete specification procedure of the CNL model. This was later proved not to hold in general by Abbé et al. (2007). Bierlaire (2006) provided a formal proof that CNL is indeed a member of GEV family of models. In contrast to a common approach of using heuristics, Bierlaire proposed to estimate the model using non-linear programming algorithms.

Wen et al. (2012) described a latent class extension to GNL. The model allows for flexible substitution patterns and estimation of market segments at the same time. It does not, however, account for the brand loyalty (or past choices) of a respondent.

The main goal of this paper is to develop a modelling approach that is computationally feasible and explicitly accounts for varying tastes and the past choice behaviour of dissimilar groups of passengers. Using a stated preference (SP) choice experiment we aim to explain differences between segments of passengers and variance in their loyalty.

This paper contributes in three ways. Firstly, we suggest a new segmentation model. Using a survey among Australians travelling to the US we explore the existence of up to eight different segments of passengers within any market. Secondly, we propose a Segment Specific Cross-Nested Logit with Brand Loyalty (SSCNL-BL) model. The model extends the CNL model by accounting for past traveller choices and membership of a particular market segment. We also test two segmentation approaches: one is a *k-means* based clustering model, and the other is a latent class derived approach. In addition, in this study we investigate how past preferences affect decisions about future airline choices. Such an approach helps deal with an inherent bias of travellers towards some travel options and leads to a more realistic and better defined model. Thirdly, the choice behaviour of Australians travelling to the United States has not yet been explored. As the routes to the United States are the second largest in terms of inbound and outbound passenger volume in the Australian market (BITRE, 2012), the results presented in this paper have the potential to contribute towards filling this gap in the literature.

The remainder of this paper is structured as follows. In the next section (Section 2) we describe our survey. Section 3 outlines the segmentation model and provides the specification of the SSCNL-BL model. In Section 4 we specify models used in this study. We then discuss our findings in Section 5 and in the last section (Section 6) we provide a summary.

2. Stated preference choice experiment

2.1. Experimental design

We utilised the *D-efficient* design to control attributes of each of the alternatives. Preliminary estimates of utility function coefficients were obtained from a small survey of students at the University of New South Wales (UNSW). Using Ngene software (Ngene, 2012) with these starting parameters we generated an efficient design to drive the questionnaire described in the next section. Using the *D-efficient* experimental design assures that, if the IIA property is violated, this violation does not arise from the errors in the experimental design. Interested readers are directed to Bliemer and Rose (2009) or Rose and Bliemer (2008, 2009) for a more detailed discussion about the *D-efficient* experimental designs.

2.2. Survey

The data for this project was collected using an online survey that was managed by a professional market research agency. A stratified sample of 360 respondents was surveyed; age, gender and geography distributions of Australian population were used as strata. Each of the respondents was asked to make two choices in three choice contexts. As a result, 720 cases for each of the three different choice contexts were collected.

The questionnaire was divided into six sections. Section 1 screened respondents to ensure they met sampling criteria such as travel behaviour and demographics. Prospective respondents who had not travelled to the United States within the three months preceding the survey were not allowed to fill in the questionnaire. Also, when age, gender or geography quotas were exhausted, potential respondents falling into any of these categories were turned away.

Section 2 of the survey probed respondents for information about their last trip to the United States – purpose of travel, options considered prior to deciding, time taken to decide and purchase, amount paid for the ticket, and frequency of travel to the United States. Respondents were also asked about purchasing channels and the number of people who travelled with them.

Sections 3–5 focused on capturing responses to choice situations. The situations were designed to track the choice behaviour of respondents throughout the booking horizon at predefined days prior to departure (DPD). Respondents were asked to consider three situations – 3 months, 1 month and 5 days before a departure. The choice situations were always presented in this order to enable the introduction of a lagged-choice component (to be discussed in detail in Section 3.2). In each choice situation we altered the operating carrier (for Qantas and United Airlines), fleet type (for Qantas), departure and flight times, price, refundability and travel class elements of the design. Respondents were asked to choose one airline product out of six. Fig. 1 shows an example of a choice scenario.

After each choice scenario we asked respondents to indicate what they would do if the option they had just selected was no longer available (see Fig. 2). Responses were captured in three categories on a 5-level scale. Respondents could also choose the 'no travel' option.

The last section (Section 6) asked respondents about their income and employment.

3. Model derivation

3.1. Segmentation model

Paraphrasing Smith (2006), Carrier (2008a, p. 7) noted that "price is the most important factor in the selection of a travel alternative, followed by flight schedule and to a much lower extent, the carrier providing the service." From a market segmentation perspective the above statement conveys important messages: (1) since there are travellers sensitive to price changes, there may also exist those for whom price is less important than other qualities, (2) since time of departure is important to some people, there may also be others for whom time of departure is insignificant, and (3) travelling with a specific carrier can be more important for some groups of passengers compared to others.

A segmentation model proposed by Belobaba (1987) accounts only for the first two factors: the price and the time of departure. However, the rapid growth of low-cost carriers during 2000s

	Qantas	United Airlines	Delta	V-Australia	Air New Zealand	Air Pacific
Operated by	Qantas	Qantas	Delta	V-Australia	Air New Zealand	Air Pacific
Fleet type	A380	B747	B777	B777	B747	B747
Departure time Arrival time	13:35 09:45	12:35 09:15	11:25 08:20	21:05 17:40	11:30 12:50	13:55 13:30
Flight time (connecting time)	13:40	13:40	13:55	13:50	18:20 (3:20)	16:35 (1:50)
Price Refundable? Travel class	2,100 No Economy	2,600 Yes Economy	2,200 No Economy	5,500 Yes Business	2,000 No Economy	1,900 No Economy
Your choice	☐	☐	☐	☐	☐	☐

Fig. 1. A sample choice scenario.

stimulated the demand for airline products. With more passengers searching for the lowest price, comparing products from various airlines on the Internet is a routine behaviour. Predictions prepared by Boeing indicate a 5% annual increase in Revenue Passenger Kilometers (RPKs) worldwide (and 4.4% for Australasia) between 2012 and 2031[3] so the competition between airlines will most likely intensify. On the other hand, frequent flyer and corporate incentive programs run by airlines are designed to encourage passenger loyalty. In such a competitive environment the sensitivity towards airlines cannot be neglected and we suggest that 'airline' should be included as the third dimension in a passenger segmentation model. Therefore, we propose the following:

Proposition 1. *In any airline market there may exist up to eight hypothetical segments of passengers whose homogeneity in behaviour can be derived from their sensitivity towards airline, price of the product and schedule.*

The segments that can be identified are defined as follows:

1. Price, time and airline **sensitive**: Passengers who are unlikely to travel if their specific requirements are not met.
2. Price and time **sensitive** but airline **insensitive**: Passengers who are unwilling to pay premium but want to travel on specific dates/time with any airline.
3. Price and airline **sensitive** but time **insensitive**: Passengers who are willing to change plans just to get the best price with the airline of their choice.
4. Price **sensitive** but time and airline **insensitive**: Passengers who are willing to switch time of departure and airline in order to secure the lowest price.
5. Price **insensitive** but time and airline **sensitive**: Business travellers willing to pay more in order to fly when and with whom they want.
6. Price and airline **insensitive** but time **sensitive**: Business passengers who travel frequently but have no specific carrier preferences.
7. Price and time **insensitive** but airline **sensitive**: Premium holiday travellers seeking the best product and willing to switch the schedule in order to secure it.
8. Price, time and airline **insensitive**: Probably the least populated group willing to trade airline and time and also willing to pay higher price.

However, we do observe that even though up to eight hypothetical clusters can be defined, some of these may not exist in some markets. This does not invalidate Proposition 1. For example, if one of the dimensions does not differentiate customers the model collapses to a simpler one with only two dimensions or even one dimension. It is also possible that only segments 5 and 4, or even only 4 are present in some markets.

3.2. Segment Specific Cross-Nested Logit model with brand loyalty (SSCNL-BL)

In this section we present two versions of our model. Firstly, let us define the utility function that is the same for both, latent-class derived and *k-means* based versions of the model.

$$U^t_{in|s} = \alpha_{i|s} + \beta^t_{i|s}\mathbf{X}^t_{in} + \beta^{t-1}_{i|s}\psi^{t-1}_{in|s} + \epsilon^t_{i|s} \tag{1}$$

We define $U^t_{in|s}$ to be the utility function for alternative i for person n who belongs to segment s at time t. The alternative specific constant, $\alpha_{i|s}$, is conditional on a market segment s in question. The $\beta^t_{i|s}$ ($\beta^{t-1}_{i|s}$) is a vector of parameters for alternative i at time t ($t-1$) and is segment s specific; we define $\beta^t_{ji|s}$ ($\beta^{t-1}_{ji|s}$) to be the jth element of the vector $\beta^t_{i|s}$ ($\beta^{t-1}_{i|s}$). The $\mathbf{X}^t_{in}$ is a vector of attributes for alternative i and socioeconomic factors of an individual n at time t, and the $\psi^{t-1}_{in|s}$ is a lagged choice response parameter, defined as:

$$\psi^{t-1}_{in|s} = \begin{cases} 1 & \text{if an alternative } i \text{ was chosen at time } t-1 \text{ by} \\ & \text{an individual } n \text{ who belongs to segment } s \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

The error terms $\epsilon^t_{i|s}$ are IID following Gumbel distribution within the market segment s.

[3] Boeing Current Market Outlook 2012 to 2031, http://www.boeing.com/assets/xls/commercial/cmo/data/CMO_2012_data.xls, accessed: April 6, 2013.

	Strongly disagree	Disagree	Not important	Agree	Strongly agree	Not travel at all
I am willing to switch airlines	☐	☐	☐	☐	☐	
I just want the lowest price	☐	☐	☐	☐	☐	
I am flexible with the departure time	☐	☐	☐	☐	☐	
						☐

Fig. 2. An example of the "what-if" section of the survey.

The unconditional probability of choosing an alternative i is calculated as follows:

$$P_n^t(i) = \sum_{s=1}^{S} P_n^t(i|s) W_n^t(s) = \sum_{s=1}^{S} \sum_{k=1}^{K} P_n^t(i|s,k) P_n^t(k|s) W_n^t(s) \quad (3)$$

where $P_n^t(i|s,k)$ is the probability of selecting an alternative i conditional on selecting nest k and being in the market segment s, $P_n^t(k|s)$ is the probability of selecting nest k conditional on belonging to the market segment s, and $W_n^t(s)$ is the level of association of an individual n to market segment s at time t.

In the latent class based version of the model, $W_n^t(s)$ is replaced by a latent probability of belonging to segment s at time t, namely $P_n^t(s)$, rendering the following form of the unconditional probability for the latent class model:

$$P_n^t(i) = \sum_{s=1}^{S} \sum_{k=1}^{K} P_n^t(i|s,k) P_n^t(k|s) P_n^t(s) \quad (4)$$

The segment probability in the latent class derived model in (4) is estimated (instead of being determined *a priori* as in the deterministic version of the model) and is defined as follows:

$$P_n(s) = \frac{e^{M_{sn}}}{\sum_{r \in S} e^{M_{rn}}} \quad (5)$$

where S is a set of all segments and the membership function M_{sn} for individual n and segment $s \in S$ has the following form:

$$M_{sn} = \gamma_{s0} + \gamma_s \mathbf{X}_{sn} \quad (6)$$

where $\mathbf{X}_{sn}$ is a set of attributes describing an individual n in segment s. The variables can be both attitudinal and behavioural. In the equation above, γ_s is a vector of segment specific coefficients and γ_{s0} is a segment specific constant.

4. Model specification

We chose 5 models to contrast their efficacy with SSCNL-BL model (with two different segmentation methods). We selected MNL as it is not capable to handle complex substitution patterns. The CNL model was chosen as it is not segment specific. An error components (EC) and a mixed logit (ML) models were picked as these are very flexible in handling correlation between alternatives. We present the EC model without explicit segmentation as this form of the model attained the highest adjusted ρ^2 score out of all the EC models tested (with and without segment specific coefficients). We also present a panel mixed logit (PML). To estimate the random coefficients for the EC, ML and PML models we used 500 draws from normal distribution. The NL model was omitted intentionally as CNL (or SSCNL-BL) collapses to NL under certain circumstances.

4.1. No market segmentation

For MNL and CNL the utility function was of the following form:

$$U_{in} = \alpha_i + \beta_{1i} \frac{\text{price}_i}{\text{income}_n} + \beta_2 \text{nref}_i + \beta_{3i} \text{business}_i + \epsilon_{in} \quad (7)$$

In the above function α_i is a constant for an alternative i, β_{1i} measures the proportion of the price of an alternative i on respondent n's income, $nref_i$ is a dummy variable for the class i refundability,[4] and $business_i$ is another dummy variable indicating whether the alternative i under consideration is a business class product. An alternative i is defined to be the operating carrier. The utility function for the EC model (Eq. (8)) includes ϵ_{in}^* that allows to specify substitution patterns. For this study the covariance matrix of the error components modelled all the possible combinations of substitutions between alternatives and were estimated together with model parameters.

$$U_{in} = \alpha_i + \beta_{1i} \frac{\text{price}_i}{\text{income}_n} + \beta_2 \text{nref}_i + \beta_{3i} \text{business}_i + \beta_{4i}^{t-1} \text{past}_{in} + \epsilon_{in}^* + \epsilon_{in} \quad (8)$$

4.2. Segmentation models

For SSCNL-BL, ML and PML we extended the utility function so the models estimate parameters for distinct market segments (indicated by *S1* and *S2* in Eq. (9)). Our segmentation model was derived from the responses to the questions presented in Fig. 2; we calculated the average value of attitudinal responses of respondents about changes in airline or departure time, and their willingness to pay a higher price. Centroids of clusters were found using a *k-means* segmentation model proposed by MacQueen (1967) and are presented in Table 1. We then assigned each respondent to the nearest segment. The numbering of segments corresponds with the definitions presented in Section 3.1.

As an outcome of an extensive *trial-and-error* phase we grouped segments {5,2,3} together into segment *S1* as this structure produced the best performing models; cluster {4} was assigned to segment *S2*. This does not negate the segmentation model

[4] Note that in all the models the coefficient for this attribute is not alternative specific but generic.

Table 1
Coordinates of centres for each segment.

Segment	Sensitivity		
	Price	Time	Airline
5	1.61	3.52	3.52
2	3.01	1.63	2.06
3	4.02	2.27	0.89
4	4.34	0.58	0.60

presented earlier as the segments are still explained by three dimensions: time, airline and price. The grouping of {5,2,3} is most likely required as a result of the small sample sizes of these individual segments.

For the latent class derivative of SSCNL-BL model we also estimated two latent segments based on a segment membership function containing the variables presented in Table 2. Passengers with higher income and who travel for business are more likely to be in the first segment. On the other hand, older travellers with higher price sensitivity have higher probability to be in the second segment. Because of these characteristics we named the segments “Wealthy and successful” and “Senior on vacation”, respectively.

4.3. Correlation over time

In addition, for all the SSCNL-BL, EC and ML models the utility function included dummy variable for the past choice, the $past_{in}$. This was set to 1 if a respondent chose the same airline in the past. For example, in the model explaining choices at 30DPD the $past_{QF}$ was set to 1 if a respondent chose Qantas at 90DPD. For the model estimated for choices at 90DPD, the $past_{in}$ variable was created using data from the revealed preference (RP) section (Section 2) of the questionnaire. Eq. (1) is then implemented in the form of Eq. (9) below with two segments and past choices included.

$$\begin{aligned} U^t_{in|s} = {} & \alpha^t_i + \beta^t_{1i|S1}\frac{\text{price}_i}{\text{income}_n} + \beta^t_{2i|S2}\frac{\text{price}_i}{\text{income}_n} + \beta^t_3\text{nref}_i \\ & + \beta^t_{4i|S1}\text{business}_i + \beta^t_{5i|S2}\text{business}_i + \beta^{t-1}_{6i|S1}\text{past}_{in} \\ & + \beta^{t-1}_{7i|S2}\text{past}_{in} + \epsilon_{in} \end{aligned} \tag{9}$$

4.4. Nesting structure

We explored two nesting structures for the SSCNL-BL models: a simple layout presented in Fig. 3, and a more complex arrangement where we tested time of departure correlations between alternatives. The latter structure performed poorly and contradicted what we have learned from Carrier (2008a). The insignificance of time of departure and flight duration in our case may be a function of two factors: (1) the length of the flight itself, and (2) the fact that we designed the experiment so the final destination was LAX. Because of the long flight duration, passengers rarely travel to the United

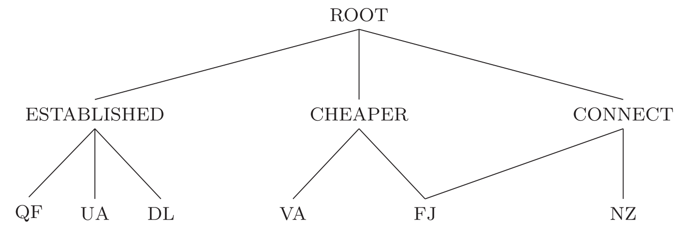

Fig. 3. CNL/SSCNL-BL – Simple nest structure.

States for one or two days, and most often stay for extended periods of time. Thus, arriving into LAX earlier or later does not factor significantly into their decision process. In addition, since the destination choice was limited to LAX, the respondents might have been more willing to take a V-Australia (VA) flight departing Sydney in late evening,[5] arriving in LAX early evening.

5. Model estimates

5.1. Performance

Assuming that all passengers behave in the same way may very likely lead to statistically insignificant estimates of their preferences; while for a certain travel product some of its attributes may be insignificant to some people, the same set of features may be the only ones that matter to another group of customers. Therefore, here we want to test the following hypothesis:

Hypothesis 1. *Extending a model to account for varying preferences of separate segments generates better results than a model that treats all passengers as a single homogeneous group.*

Table 3 summarises the results of all the models.[6] As expected, the SSCNL-BL model that utilizes deterministic segmentation performs similarly to the more complex ML. The better results of SSCNL-BL models are not just an outcome of more variables being included in the utility function. The SSCNL-BL models clearly outperforms MNL, CNL or EC. The latent class version of SSCNL-BL (SSCNL-BLL) model outperforms ML in every choice context. It proves that estimating independent price coefficients for separate segments and adjusting for the past choice has a considerable influence on the final result, and hence corroborates our Hypothesis 1. For these heterogeneous groups of passengers our model estimates price coefficients that are statistically significant (refer to Table 4, page 6 and Table 6, page 7).

What comes as a surprise is the very similar performance of the CNL model when compared to MNL. Still, CNL allows for a better understanding of the competitive forces among alternatives grouped in the same nest (Garrow, 2010). The values of logsum parameters θ_k and nest membership η_{ik} for SSCNL-BL are summarized in Table 5, and for SSCNL-BLL in Table 7 (page 12). We noted high levels of competition in the “Established” nest; correlation between QF, UA and DL equals 51%.[7] Only for 5DPD did “Cheaper” and “Connect” nests show increased competition; the “Cheaper” nest showed even greater correlation than the “Established” nest – 56%. This is most likely due to the proximity to the departure date –

Table 2
Coefficients for latent classes in the latent derived SSCNL-BL model.

Parameter	Segment	5DPD[a,b]	30DPD	90DPD
Constant	Wealthy and successful	3.74**	6.45**	5.46**
Travel for business		0.82*	1.13**	
Income		0.76	1.40*	1.16*
Age	Senior on vacation	0.32	0.84	1.20**
Price sensitivity		5.77**	9.60**	7.58**

[a] * – Statistically significant at 10% level of confidence.
[b] ** – Statistically significant at 5% level of confidence.

[5] At the time of designing the SP experiment V-Australia offered a late night flight to the U.S.

[6] All the estimations were performed on a MacBook Pro running Mac OS X Mountain Lion using Intel Core i7 quad-core processor clocked at 2.6 GHz and 16 GB of RAM. We used Python Biogeme (Bierlaire, 2003) that utilized 8 threads to estimate the models.

[7] SSCNL-BL, 5DPD model. The correlation is calculated using the following formula: $r = (1 - \theta_k^2)$, see Garrow (2010).

Table 3
Comparison of models efficiency.

Measure	DPD[a]	MNL	CNL	SSCNL[b] BL	SSCNL[c] BLL	EC[d]	ML	ML panel
Final LL	90DPD	−1119	−1118	−1055	−1032	−1097	−1037	−2988
	30DPD	−1064	−1061	−968	−923	−1021	−947	
	5DPD	−1070	−1066	−994	−958	−1079	−978	
LL test	90DPD	342.07	343.67	470.43	515.07	386.95	504.17	1762.97
	30DPD	451.31	457.86	642.23	732.24	537.62	684.92	
	5DPD	439.05	446.62	590.92	664.05	420.62	624.12	
ρ^2	90DPD	0.133	0.133	0.182	0.200	0.153	0.195	0.230
	30DPD	0.175	0.177	0.249	0.284	0.241	0.265	
	5DPD	0.170	0.173	0.229	0.257	0.217	0.242	
Adjusted ρ^2	90DPD	0.122	0.118	0.158	0.174	0.133	0.165	0.220
	30DPD	0.164	0.163	0.224	0.257	0.222	0.235	
	5DPD	0.159	0.158	0.204	0.231	0.198	0.212	
Est. time (m:s)	90DPD	0:00	0:06	0:05	0:41	02:29	1:58	
	30DPD	0:00	0:03	0:06	0:50	03:12	2:27	38:33:00
	5DPD	0:00	0:04	0:07	0:59	02:47	1:25	

[a] Days prior to departure.
[b] *k-means* segmentation.
[c] Latent class.
[d] Error components model.

if a passenger needs to travel on a specific date, he/she examines the alternatives in greater depth. Conversely, we observed almost no competition in "Cheaper" and "Connect" nests for 30DPD and 90DPD. The value of 0.23 for the "Cheaper" nest at 90DPD may indicate an extremely high correlation but this nest has only one alternative, showing that the model is no longer a CNL but NL. Thus, the low value of this coefficient makes the VA option less likely to be chosen. This also means that if one of the alternatives was not available the probability of choosing another within the nest would not change. For the SSCNL-BLL the nesting structure reveals the highest degree of competition for the "Established" nest at 30DPD. At 5DPD the "Connecting" nests shows more competition than any other but FJ barely reveals any connection to that nest.

Table 4
Parameters estimated, SSCNL-BL.

Parameter	Airline	Segment	5DPD[a,b]	30DPD	90DPD
Constant	DL	–	−0.53*	−0.5	−0.43
Constant	FJ	–	−1.31**	0.01	−0.54
Constant	NZ	–	−0.32	0.35	−0.32
Constant	QF	–	0.00 fixed	0.00 fixed	0.00 fixed
Constant	UA	–	−0.32	0.18	−0.27
Constant	VA	–	0.64**	1.37**	0.46
Business	All	{5,2,3}	−1.54**	−1.39**	−0.72*
Business	All	{4}	−2.60**	−1.20**	−1.80**
Non refundable	All	–	0.48**	0.70**	0.63**
Price/income	DL	{5,2,3}	−1.13**	−1.80**	−0.82**
Price/income	FJ	{5,2,3}	−1.90**	−1.21**	−0.80**
Price/income	NZ	{5,2,3}	−1.10**	−1.17**	−0.59**
Price/income	QF	{5,2,3}	−0.98**	−1.10**	−0.83**
Price/income	UA	{5,2,3}	−0.97**	−1.17**	−0.69**
Price/income	VA	{5,2,3}	−1.28**	−1.11**	−0.81**
Price/income	DL	{4}	−1.62**	−2.10**	−1.74**
Price/income	FJ	{4}	−1.71**	−1.40**	−1.59**
Price/income	NZ	{4}	−1.66**	−1.43**	−1.67**
Price/income	QF	{4}	−1.46**	−1.17**	−1.38**
Price/income	UA	{4}	−1.85**	−1.31**	−1.52**
Price/income	VA	{4}	−1.91**	−1.69**	−1.68**
Past choice	FJ	–	1.21**	1.47**	−1.30
Past choice	NZ	–	0.72**	0.81**	0.95*
Past choice	QF	{5,2,3}	1.62**	2.61**	1.98**
Past choice	QF	{4}	2.50**	1.71**	0.44
Past choice	UA	–	1.16**	1.18**	0.33
Past choice	VA	{5,2,3}	1.23**	0.59*	1.21**
Past choice	VA	{4}	1.30**	0.86**	0.55

[a] * – Statistically significant at 10% level of confidence.
[b] ** – Statistically significant at 5% level of confidence.

5.2. Segment shifts

In our survey we asked passengers to rate their willingness to pay a higher price, and their attitude towards changing an airline and/or departure time should their preferred alternative become unavailable. Their responses were captured on a five level scale. Our experiment also set three different choice contexts for the respondents: 90DPD, 30DPD and 5DPD. This allowed us to track shifts in respondents' attitudes throughout the booking horizon. We designed our survey in this way to test the following hypothesis:

Hypothesis 2. *Closer to the departure day customers become more price insensitive, and airline and time sensitive.*

Fig. 4 illustrates the shifts between segments throughout the booking horizon. We observed that the number of respondents in the most price insensitive segments (5) and (2) increased in time from 98 passengers at 90DPD to 137 passengers at 5DPD. Consequently, segments (4) and (3) reduce in number the closer to the departure date. The number of passengers for these two segments combined shrunk from 262 at 90DPD to 223 at 5DPD. These findings confirm the Hypothesis 2.

5.3. Willingness-to-pay

In addition, because only Australians took part in our research, we also expected that:

Hypothesis 3. *There exists a stronger preference towards Qantas among Australians travelling to the United States.*

We noticed passengers' willingness-to-pay premium shifts with time at different paces for the two groups of segments we defined

Table 5
Logsum parameters θ_k and nest membership η_{ik}, SSCNL-BL.

Parameter	Airline	5DPD	30DPD	90DPD
Cheaper		0.66**	1.00**	0.23**
Connecting		0.81**	1.00**	0.89**
Established		0.70**	0.62**	0.68**
Cheaper	FJ	0.46	0.07	0.00
Cheaper	VA	1.00 fixed	1.00 fixed	1.00 fixed
Connecting	FJ	0.54	0.93	1.00
Connecting	NZ	1.00 fixed	1.00 fixed	1.00 fixed
Established	DL	1.00 fixed	1.00 fixed	1.00 fixed
Established	QF	1.00 fixed	1.00 fixed	1.00 fixed
Established	UA	1.00 fixed	1.00 fixed	1.00 fixed

Table 6
Parameters estimated, SSCNL-BLL.

Parameter	Airline	Segment	5DPD[a,b]	30DPD	90DPD
Constant	DL	–	−1.45**	−2.17**	−1.33**
Constant	FJ	–	−2.45**	−1.73**	−1.35**
Constant	NZ	–	−1.07**	−1.15**	−0.86**
Constant	QF	–	−0.66*	−2.07**	−0.87**
Constant	UA	–	−0.92**	−1.48**	−0.86**
Constant	VA	–	0.00 fixed	0.00 fixed	0.00 fixed
Non refundable	–	–	0.45**	0.68**	0.53**
Business class	–	–	0.82*	1.13**	0.00**
Price/income	DL	Young	−0.82**	−1.23**	−0.51**
Price/income	FJ	Young	−0.91**	−0.93**	−0.40**
Price/income	NZ	Young	−0.87**	−0.92**	−0.39**
Price/income	QF	Young	−0.79**	−0.72**	−0.45**
Price/income	UA	Young	−0.88**	−0.94**	−0.49**
Price/income	VA	Young	−1.01**	−0.97**	−0.47**
Price/income	DL	Seniors	−9.10**	−10.00**	−9.31**
Price/income	FJ	Seniors	−8.76**	−10.30**	−8.58**
Price/income	NZ	Seniors	−9.45**	−13.00**	−10.00**
Price/income	QF	Seniors	−8.68**	−9.67**	−8.00**
Price/income	UA	Seniors	−9.34**	−13.80**	−8.66**
Price/income	VA	Seniors	−8.85**	−20.00**	−7.90**
Past choice	DL	–	−0.46	1.42**	0.00
Past choice	FJ	–	1.33**	1.56**	0.08
Past choice	NZ	–	0.95**	0.88**	0.99*
Past choice	QF	–	1.92**	2.64**	1.50**
Past choice	UA	–	1.16**	1.64**	0.36
Past choice	VA	–	1.17**	0.71**	1.03**

[a] * – Statistically significant at 10% level of confidence.
[b] ** – Statistically significant at 5% level of confidence.

earlier ({5,2,3} and {4}). Fig. 5 shows passengers' willingness-to-pay (WTP) for a business class at 90DPD, 30DPD and 5DPD. These charts confirm our earlier findings that significant differences exist between segments. In addition, respondents seem to be favouring QF over all the other airlines as they are willing to pay more for its business class, irrespective of the segment to which they belong except for 90DPD and 5DPD scenarios and segments {5,2,3}. As expected, for these segments we observe higher WTP but the differences in levels are not as pronounced as in segment {4}. This verifies our Hypothesis 3 and also supports our segmentation model: passengers in segments {5,2,3} are more loyal to the airline of their choice.

Strong competition can be observed between the three airlines grouped under the "Established" nest: QF, UA and DL. The popularity of QF business class may be credited to a couple of factors: (1) QF is the biggest and the oldest of Australia's carriers with a history of no major accidents, and (2) use of A380s and higher standards in the business and first class cabins. However, we found that the type of aircraft played no major role in the passenger decision process. Respondents were also willing to pay substantially higher fares to travel business class with DL. This is a factor of DL's pricing that was slightly lower than QF and UA, and a newer B777s operating the

Table 7
Logsum parameters θ_k and nest membership η_{ik}, SSCNL-BLL.

Parameter	Airline	5DPD	30DPD	90DPD
Cheaper		0.78**	1.00**	1.00
Connecting		0.53	0.75	1.00
Established		0.65**	0.48**	0.65**
Cheaper	FJ	0.94**	0.93**	0.04**
Cheaper	VA	1.00 fixed	1.00 fixed	1.00 fixed
Connecting	FJ	0.06**	0.07**	0.96**
Connecting	NZ	1.00 fixed	1.00 fixed	1.00 fixed
Established	DL	1.00 fixed	1.00 fixed	1.00 fixed
Established	QF	1.00 fixed	1.00 fixed	1.00 fixed
Established	UA	1.00 fixed	1.00 fixed	1.00 fixed

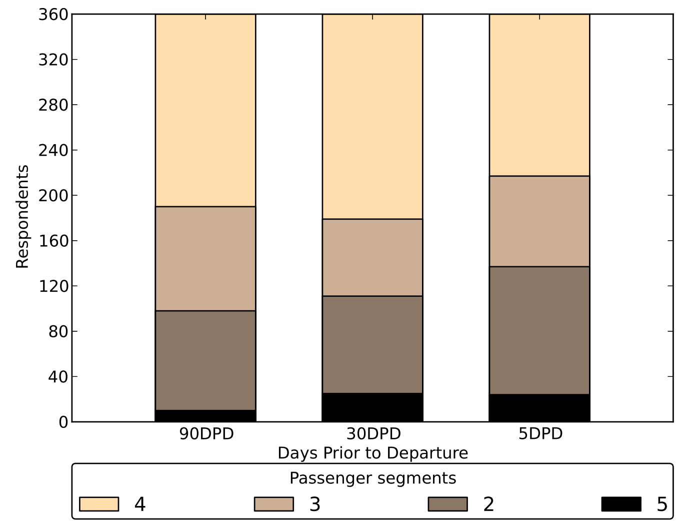

Fig. 4. Segment distribution throughout booking horizon.

route when compared to B747s used by UA. UA's popularity is most likely historical and an effect of slightly lower pricing when compared to QF.

5.4. Brand loyalty

The specific design of the questionnaire that allowed us to track passengers' choices throughout the booking horizon also allowed us to measure influence of the past choices (or brand loyalty) on their present choice. We intentionally chose such a design to test the following hypothesis:

Hypothesis 4. *Brand loyalty has a positive impact on the passenger choice*.

The results presented in Tables 4 and 6 confirm our Hypothesis 4. For the two SSCNL-BL models, all the 'Past choice' coefficients but one are positive and most are highly statistically significant. The FJ lagged choice parameter for the 90DPD SSCNL-BL model is negative but not statistically significant. Also, the past choice coefficient for DL in the SSCNL-BLL model is negative but insignificant. Hence, these may be removed from the utility functions of these two airlines without loosing any accuracy. We noted a stronger respondent preference for QF; at any point throughout the booking horizon, respondents who had travelled with QF in the past are less likely to switch to other airline. A smaller value of the coefficient for the group of segments {5,2,3} in the 5DPD model means that the respondents examined other options in greater detail, and brand loyalty is not as stray as in segment {4}.

5.5. Computational efficiency

The last hypothesis we wanted to test deals with the speed of estimation of SSCNL-BL models and their efficiency when compared with other models. It can be succinctly expressed as below:

Hypothesis 5. *Estimating SSCNL-BL takes significantly less time than estimating ML while attaining similar efficiency.*

We stated earlier that the performance of our models is similar to the ML model without the computational burden of estimating a model with a utility function that does not have a closed form. We expected that would be the case when we were designing the SP experiment and the SSCNL-BL model.

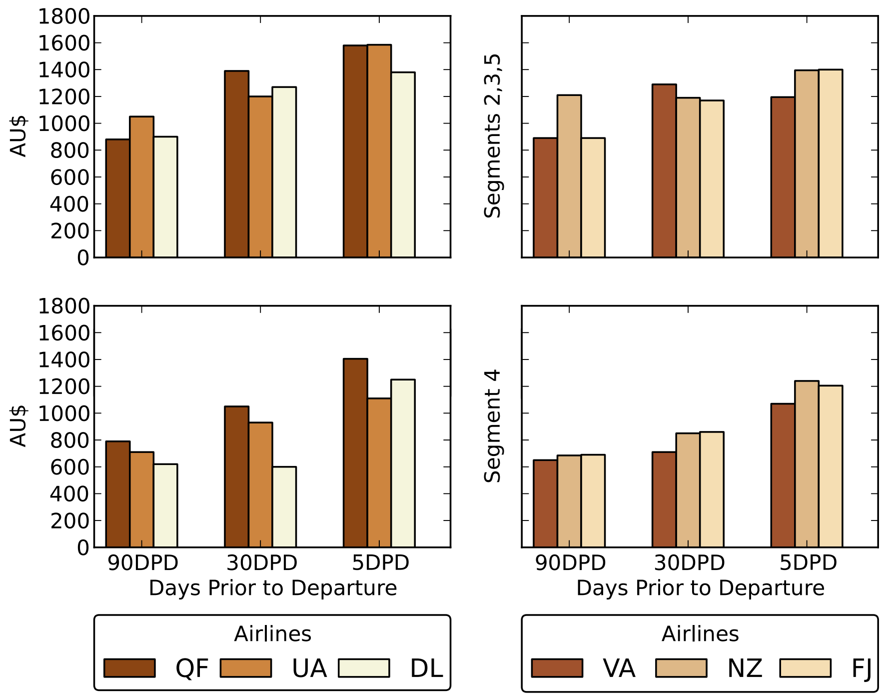

Fig. 5. Willingness-to-pay (WTP) by segments and DPD.

SSCNL-BL is the fastest to estimate when compared with either SSCNL-BLL, EC, ML, not to mention PML. The latent class based SSCNL-BLL produced the best results in terms of the adjusted ρ^2 but this came at the cost of a longer estimation time. In any choice context, the EC and ML models have always been the slowest. It has to be stressed that SSCNL-BLL delivered better results than ML in every single choice context. Therefore, when performance is required (i.e. with extensive projects) a researcher can choose SSCNL-BL that is more accurate than MNL at a very slight increase in estimation time. However, if accuracy is more important then SSCNL-BLL is a viable option as it renders better results than ML or PML and is also faster. We leave out for future research a task of comparing SSCNL-BL models' sensitivity to sample size. It is expected that with increasing sample size the gap in estimation time of SSCNL-BL models when compared with ML will become wider. Nevertheless, in contrast to models with a closed-form utility function (like MNL, CNL or SSCNL-BL), ML, EC or PML help to better understand the distribution of passengers' preferences; where such knowledge is required these models would be a better fit.

5.6. Implications for the industry

For many years, discrete choice models (DCMs) have been proposed to solve airline problems such as allocating seats while considering customer choices. Examples include works of Talluri and van Ryzin (2004) or Gallego et al. (2009, 2007). The most prevalent model used in such applications was the Multinomial Logit model due to its ease of estimation and adaptation. However, as we have shown, the deterministic form of SSCNL-BL clearly outperforms MNL with a marginal increase in computational load.

Carrier (2008a) mentioned that his model did not allow for substitution patterns. Our model addressed this problem. An airline analyst using SSCNL-BL has the tools to quantify competitive forces between airlines, and can then apply this knowledge when making pricing and seat allocation decisions. For example, in market with four hypothetical airlines, where the fiercest competition is between two of them, SSCNL-BL can identify the direct competitor. The analyst can focus Revenue Management efforts that target this particular airline. Of course, she should not neglect other airlines completely but given the quantity of available data, the airline employee is at risk of losing focus. Instead of analysing all of the data, the analyst can focus on the most important 20% that brings 80% of the revenue. Our model enables prioritization of work and increases effectiveness.

DCMs can also be employed to estimate upsell or cross-sell behaviour of passengers. The model presented in this paper simultaneously estimates price coefficients of various distinct segments, giving the analyst a deeper understanding of the market demand. For example, with in-depth market knowledge, an analyst can safely assume the probability of upsell is very small for the segment that is airline and time insensitive, and price sensitive. Actions towards these passengers can then focus on trying to market seats on other flights offered by the airline.

Applications of DCMs to Revenue Management are not limited to seat allocation models. Ratliff et al. (2008) successfully applied MNL to estimate true demand, spill and recapture in a multi-airline, multi-flight and multi-class environment. Our model can help to estimate more robust spill and recapture rates as it accounts for inter-airline and inter-product substitution patterns. This results in better estimates of the true demand and leads to more robust forecasts of future true demand, limiting the spiral down effect described by Boyd et al. (2001).

The main limitation of our study comes from the use of SP data. SP data are useful when estimating or testing discrete choice models or when an airline wants to learn more about one of the markets and/or their passengers in hypothetical scenarios. However, practical applications of the SP data are limited due to their static nature – it takes time and money each time a researcher or an airline wants to gather such data. On the other hand, relying purely on the RP data (e.g. booking data) leads to a myopic view of the market because RP data (as its name suggests) only reveal the choices made, not the choice-making behaviour. In addition, RP data can be highly collinear (Louviere et al., 2000) and that further raises concerns about their usability. Recreating the passenger choice set in a multi-airline, multi-flight and multi-class environment is not a trivial task with only RP data. Carrier (2008b) proposed an approach where he combined CRS and frequent flyer data to recreate passenger choices. The SP experiment presented in this paper emphasizes the intricacies of inter-airline interactions resulting from the complexity of passengers behaviour. This, in return, may help in understanding that the unconditional reliance on the historical booking data leads to decisions being made with a very limited knowledge. In the long term, strategies stemming from such a decision-making process may be detrimental to an airline revenue.

6. Summary

In this paper we presented a novel segmentation model that distinguishes passengers by airline, time and price substitutions behaviour. Thanks to the proliferation of the Internet, cheap and easy ways of comparing alternatives add to the complexity of modelling customers behaviour. The presented approach allows airlines to differentiate customers in a more robust way. We observed that even though it is possible to identify up to eight theoretical segments, in some markets only three or four will be identifiable.

We then exploited the segmentation model by developing the Segment Specific Cross-Nested Logit with Brand Loyalty (SSCNL-BL) model. The model allows an analyst to jointly estimate separate coefficients for distinct market segments, and account for past passenger preferences and/or choices. We also proposed a latent class extension to the model rendering better defined segments. We compared the SSCNL-BL with 5 different discrete choice models. Results show that the deterministic version of SSCNL-BL model outperforms MNL and CNL arriving at similar results as the more complex ML model without requiring extensive computations. The probabilistic (latent) form of the model (SSCNL-BLL model) outperforms ML in every choice context. Our study also shows that Australian passengers travelling to the United States prefer Qantas as their main carrier.

References

Abbé, E., Bierlaire, M., Toledo, T., 2007. Normalization and correlation of cross-nested logit models. Transportation Research Part B: Methodological 41 (7), 795–808.

Belobaba, P.P., May 1987. Air Travel Demand and Airline Seat Inventory. Ph.D. thesis. Massachusetts Institute of Technology.

Ben Akiva, M., Lerman, S.R., 1985. Discrete Choice Analysis: Theory and Application to Travel Demand. The MIT Press.

Bierlaire, M., 2003. BIOGEME: a free package for the estimation of discrete choice models. In: Proceedings of the 3rd Swiss Transportation Research Conference, Ascona, Switzerland. URL. http://biogeme.epfl.ch.

Bierlaire, M., 2006. A theoretical analysis of the cross-nested logit model. Annals of Operations Research 144 (1), 287–300.

BITRE, 2012. Bureau of Infrastructure, Transport & Regional Economics. In: International Airline Activity – Time Series. http://www.bitre.gov.au/publications/ongoing/international_airline_activity-time_series.aspx (last accessed 04.05.12.).

Bliemer, M.C.J., Rose, J.M., 2009. Designing stated choice experiments: state-of-the-art. In: Kitamura, R., Yoshii, T., Yamamoto, T. (Eds.), The Expanding Sphere of Travel Behavious Research. Emerald, pp. 499–537.

Boyd, E.A., Kambour, E., Tama, J., 2001. The impact of buy-down on sell up, unconstraining, spiral-down. In: 1st Annual INFORMS Revenue Management Section Conference. INFORMS, New York.

Carrier, E., 2008a. Modelling the Choice of an Airline Itinerary, Fare Product Using Booking, Seat Availability Data. Submission to Anna Valicek 2008 Prize. AGIFORS.

Carrier, E., 2008b. Modelling the Choice of an Airline Itinerary and Fare Product Using Booking and Seat Availability Data. Ph.D. thesis. Massachusetts Institute of Technology, Department of Aeronautics and Astronautics.

Gallego, G., Li, L., Ratliff, R.M., May 15 2007. In: Revenue Management with Customer Choice Models. AGIFORS RMD, Cargo Study Group, South Korea.

Gallego, G., Li, L., Ratliff, R.M., 2009. Choice-based EMSR methods for single-leg revenue management with demand dependencies. Journal of Revenue & Pricing Management 8 (2/3), 207–240.

Garrow, L., 2010. Discrete Choice Modelling, Air Travel Demand – Theory, Applications. Ashgate Publishing Limited.

Greene, W.H., Hensher, D.A., 2003. A latent class model for discrete choice analysis: contrasts with mixed logit. Transportation Research Part B: Methodological 37 (8), 681–698.

Hess, S., 2007. Posterior analysis of random taste coefficients in air travel behaviour modelling. Journal of Air Transport Management 4 (13).

Louviere, J., Hensher, D.A., Swait, J.D., 2000. Stated choice methods: analysis and application. University Press, Cambridge.

MacQueen, J., 1967. Some methods for classification, analysis of multivariate observations. In: Proceedings of Fifth Berkeley Symposium on Mathematical Statistics, Probability. University of California Press, pp. 281–297.

McFadden, D.L., 1974. Frontiers in Economics. In: Conditional Logit Analysis of Qualitative Choice Behaviour. Academic Press, New York, Ch, pp. 105–142.

McFadden, D.L., 1978. Modelling the choice of residential location. In: Spatial Interaction Theory, Planning Models. North Holland, Amsterdam.

Ngene, Feb 16 2012. Ngene 1.1.1 – The Cutting Edge in Experimental Design. ChoiceMetrics. www.choice-metrics.com.

Papola, A., 2004. Some developments on the cross-nested logit model. Transportation Research Part B: Methodological 38 (9), 833–851.

Ratliff, R.M., Rao, B.V., Narayan, C.P., Yellepeddi, K., 2008. A multi-flight recapture heuristic for estimating unconstrained demand from airline bookings. Journal of Revenue & Pricing Management 7 (2), 153–171.

Rose, J.M., Bliemer, M.C.J., 2008. Handbook of Transport Modelling. In: Stated Preference Experimental Design Strategies. Elsevier, Oxford. Ch, pp. 151–180.

Rose, J.M., Bliemer, M.C.J., 2009. Constructing efficient stated choice designs. Transport Reviews 29 (5), 587–617.

Small, K., 1987. A discrete choice model for ordered alternatives. Econometrica 55 (2), 409–424.

Smith, B., 2006. Helping Travelocity Transition to Travel Retailer. Presentation to the MIT Operations Research Seminar, after Carrier (2008a).

Talluri, K.T., van Ryzin, G.J., 2004. Revenue management under a general discrete choice model of consumer behaviour. Management Science 50 (1), 15–33.

Train, K., 2009. Discrete Choice Methods with Simulation, second ed. Cambridge University Press.

Vovsha, P., 1997. Cross-Nested Logit Model: an Application to Mode Choice in the Tel-Aviv Metropolitan Area. Transportation Research Board 76th Annual Meeting.

Wen, C.-H., Huang, W.-W., Fu, C., Chou, P.-Y., 2012. A latent class generalised nested logit model and its application to modelling carrier choice with market segmentation. Transportmetrica, 1–20.

Wen, C.-H., Koppelman, F.S., 2001. The generalized nested logit model. Transportation Research Part B: Methodological 35 (7), 627–641.