Passenger segments and their fare class choice behaviour on competitive routes in south China

Jue Wang 1 , Brett Smith 2 , Min Qiu 3

^{1,2,3}The University of Western Australia, Crawley, 6009, WA, Australia Email for correspondence: Brett.Smith@uwa.edu.au

ABSTRACT

In this research, we developed a methodology to analyse traveller's fare product choice in a competitive market using stated choice data. To maintain the non-overlapping hierarchical price structure applied in revenue management system, an availability design was proposed, in which the 'presence' and 'absence' of alternatives in each choice set is controlled by an orthogonal fraction of a 2^J factorial design.

The estimation results of a random parameter model with presence and absence effects showed that travellers who book their tickets within one week are less price-sensitive to those booking longer ahead. In addition low income travellers are much more price-sensitive.

Keywords: Airline Revenue Management, Demand Dependency, Discrete Choice Modelling, State Preference Data, Experimental Design

1 INTRODUCTION

In the early regulated aviation markets full-serviced first class and unrestricted coach fares were tied to the mileage-based fare calculations. Deregulation meant that service-based pricing was replaced by product differentiation strategy using both service and ticket conditions. Under a typical differentiated fare structure, each fare product is designed for a specific demand segment with distinct level of willingness-to-pay (WTP). At the high end, airlines offer premium fares with enhanced service amenities and fewer restrictions, and at the low end, discount fares with more rigorous fare rules are often applied (Belobaba, 2009). To avoid the diversion of high WTP customers, the availability of discount fares are often controlled by revenue management systems (RMSs).

Traditional SP designs applied to air ticket choice do not maintain the dependence of air fare on the booking class¹. This is because most of the design methods include all booking classes. For such designs, the hierarchical fare structure is very likely to cause dominant choices of the discount fares. This is especially true when mild fare restrictions are present. The price levels need to overlap the ticket classes (i.e. a more restrictive fare could be presented as the cheaper option) to force trade-offs. The parameter estimates generated in this way are only valid for the stated choice scenarios rather than a true reflection of market behaviour. To mimic actual market behaviour a few studies have included a search mechanism (Proussaloglou and Koppelman (1999), Collins et al. (2010) and Bliemer and Rose (2011). This study includes a presence/absence or an *availability* design to imitate the market when ticket classes are closed due to prior purchases.

The availability design maintains the hierarchical fare structure applied in the airline market by presenting choice sets with a systematically controlled 'availability' of fares. The estimation of an error components model shows that such a design scheme is able to capture travellers' preference heterogeneity. The strategy offers a potentially more revealing and reliable method to estimating passengers' choice of air tickets than design methods that assume fixed number of alternatives in the choice sets.

The rest of the paper is organised as follow. The next section introduces the experimental design method used in this research. This is followed by a presentation of the empirical setting and the estimation results of an empirical study conducted on Chinese domestic airline market. The paper is closed in section 4 with a conclusion.

2 EXPERIMENTAL DESIGN

 \overline{a}

Experimental design is the process of generating a matrix of values to allocate attribute levels that describe hypothetical options in a stated choice survey. A good design allows analysts to detect the influence of the attributes upon observed choices made by sampled respondents (Rose & Bliemer, 2009). Orthogonal and balanced incomplete block designs represent one strategy to isolate the effect of an attribute (Bunch et al., 1996; Louviere et al., 2000). However, these designs can often confound the primary attribute effects with higher order interaction effects. Full factorial designs eliminate this confounding but they are rarely practical, especially in choice settings where many attributes and their interactions affect choice. In addition, orthogonality is sensitive to practical survey concerns, such as the inclusion of socio-demographic variables in the estimation model, missing data and sampling issues (Rose & Bliemer, 2004).

An alternate design strategy is to minimise the error (variance) introduced in any partial design. Efficient designs aim to produce stable and reliable parameter estimates in a fractional design setting (Ryan et al., 2007). This is achieved by minimising at least one property of the asymptotic variance–covariance (AVC) matrix: determinant, trace, variances. The smaller the standard errors introduced by the design are for the parameter estimates,

 $¹$ Notably, booking class should be differentiated from anther concept, the travel class (e.g. business class,</sup> economy class). Normally, each travel class may contains several booking classes, each of which is fenced by certain fare rules and price level, and provided with a unique code.

the more likely the statistical inferences are to be correct. The non-linear functional form of choice models and the fact that the parameter estimates are determined by attribute differences cause added complications in determining an efficient design. Namely, the combination of attribute levels presented to the respondent and the 'true' parameter of choice directly affect the standard errors.

McFadden (1974) derived the AVC matrix for the multinomial logit model by solving the second derivatives of its log-likelihood function with respect to the parameters. The same method is used to arrive at the AVC matrix for other choice model error structures (see Huber and Zwerina, 1996; Kanninen, 2002; Sándor and Wedel, 2002; Carlsson and Martinsson, 2003; Ferrini and Scarpa, 2007; Bliemer et al., 2009a,b; Rose and Bliemer, 2009). One of the challenges present at the design stage is the requirement for a known or previously estimated parameter vector when deriving the AVC. Obtaining reliable prior estimates aids the design process, because a misspecification leads to losses in efficiency (Kessels et al., 2008; Bliemer et al., 2009b). Priors may be obtained from literature reviews or pilot studies. However, in relatively new markets the researcher may be left with no other choice than to use null-parameter priors.

The empirical case presented here is highly specific $-$ a short haul domestic flight departing from Shanghai International airport. In addition, this is the first availability design for air ticket choices that we are aware of. For these two reasons, we applied a D_z -error criteria, which assumes a value of zero for the prior parameter estimates. The alternatives in the design were defined as all the possible combinations of fare products, itineraries, and carriers in a target origin-destination market. Since this research focused on domestic travel, an itinerary was defined as a non-stop flight departing from Shanghai International to nearby airports with flight durations less than 120 minutes. The fare products were characterised by price, departure time, on board amenities and fare class restrictions.

Non-overlapping hierarchical levels of price were assigned to each fare product given its booking class. To avoid dominant choices, the presence of each alternative in a choice set is controlled by an orthogonal fraction of a 2^{J} factorial design. The advantage of such a practice is that the dependency between alternatives perceived by decision makers can be estimated independently from the co-occurrence of alternatives. The design follows the two-step sequential 'availability design' proposed by (Louviere et al., 2000). However, their method imposed orthogonality on both the availability of alternatives and attribute levels within each alternative. Such a design in our setting would require a very large number of respondents or place an extremely high burden on the response task for a reasonable sample size. In comparison, the 'efficient' availability design applied in this research contained much fewer choice scenarios. To generate the design, the following model was used.

$$
V_{i|A} = \alpha_i + \beta X_i + \sum_{i' \neq i} Z_{i'} \delta_{ii'}
$$

$$
\begin{array}{c}\n(1)\n\end{array}
$$

where

 α_i = intercept for fare class *i*, *i*=1,...,*M*

 β = price effect,

= the price levels of the ith fare class,

 $\delta u'$ = the availability cross effect of fare class v' on fare class i,

 $\overline{Z}_{i'} = 1$ if fare class i' presents in the choice set A, -1 otherwise.

The above model takes a standard multinomial logit form. The Log-likelihood function is:

$$
L(\beta | x, y) = \sum_{n=1}^{N} \sum_{j=1}^{m} y_{j_{n}n} log P_{j_{n}n}(x_{n} | \beta)
$$
\n(2)

where y describes the outcomes of all choice tasks, that is, $y_{f_n n}$ equals to one if alternative f_n is chosen in the choice set n , and zero otherwise. The AVC matrix can be derived from the second derivative of the log-likelihood function. Since the cross-alternative parameters are alternative specific, the elements in the matrix are based on the specification in Bliemer and Rose (2005). The availability design permits a violation of the regularity condition of random utility models. If the parameter estimate

Notably, with the existence of cross-alternative attributes, the IIA property of the MNL no longer holds. Such a model is no longer a random utility model since the utility of one alternative depends on the attributes of other alternatives (Brownstone and Train, 1999). On the contrary, it is more in line with the so-called 'mother logit' model (McFadden, 1975) which is a standard logit model with attributes of other alternatives entered in the representative utilities. For studies of similar models please refer to Batsell and Polking (1985), Jain and Bass (1989), Timmermans and Borgers (1991), and Lazari and Anderson (1994).

3 EMPIRICAL INQUIRY

A paper and pen survey was administered to travellers on business and non-business trips over the period of $15th$ March to $5th$ April 2012 in Shanghai, China. A total of 485 questionnaires were collected, among which 305 were collected in the departure lounge at Shanghai Hongqiao International Airport. The rest were completed by travellers who booked tickets at a local travel agent. All together the exercise produced 3734 choice observations. This paper is limited to the choice behaviour of the 326 passengers travelling for personal reasons.

3.1 Questionnaire Design

The first part of the questionnaire includes general information of travellers and trip related information regarding their travel experience within the last 12 months. Other than purpose of travel respondents indicated the number of days that they had purchased their ticket prior to departure, how frequently they fly, their personal income and rating for the importance of inflight quality indicators.

The second part contains scenarios from a stated choice experiment generated using the method proposed in the last section. In each scenario, respondents were asked to select a ticket for a short haul trip departing from Shanghai. Figure 1 is an example of the choice scenarios used in the questionnaire.

					No.1
Carrier	Flight	Departure	Premium Eco	Semi Flex	None Flex
Major		Morning	¥600	¥450	¥400
Low cost	2	Morning			Sold Out
Major	3	Midday	Sold Out	Sold Out	Sold Out
Major	$\overline{4}$	Afternoon	Sold Out	Sold Out	Sold Out
Major	5	Night	¥550	Y500	¥350
Low cost	6	Night	\sim	$\overline{}$	¥300
					Not Travel \Box

Figure 1 A Scenario in the Stated Choice Experimental Design

As shown in Figure 1, there are two carriers in this hypothetical market. The major carrier offered four flights a day with three distinct booking classes, and the low cost carrier offered a single ticket class for its two flights per day. Each class was assigned a given fare rule varying from 'no change can be made' (i.e. non-refundable and non-transferable) to 'fully flexible' (i.e. the ticket is valid for 12 months and redeemable against any booking without penalty for changes to the itinerary). A non-overlapping price schedule is used as presented in Table 1. Notably, in order to distinguish the non-flexible fare products provided by the two different carriers, the price of the low-cost fares was set to be slightly lower than those offered by their major opponent.

The fare classes are offered at four time slots throughout the day. The 8:00am morning and 8:00pm evening flights were early and late enough to signal an inconvenience. Alternatively the two day flights, whilst being more convenient for accessing and egressing the airports, disrupted the working day.

Booking Class	Fare Rules and Price Range	
	No change can be made. No refund will be provided when	
Non-flexible Economy (Low cost)	cancellation occurs.	
	Price level: ¥300, ¥350	
	No change can be made. No refund will be provided when	
Non-flexible Economy (Major)	cancellation occurs.	
	Price levels: $Y350, Y400$	
	Change will generate a fee equal to 30% of the ticket price.	
	Cancellation will generate a fee equal to 50% of the ticket	
Semi-flexible Economy (Major)	price.	
	Price levels: $Y450, Y500$	
	No charge will be made when change occurs. Cancellation	
Premium Economy (Major)	will generate a fee equal to 5% of the ticket price.	
	Price levels: $Y550, Y600$	
Choose not to book	Reference Alternative	

Table 1 – Fare rules and Price of the Booking classes

The design produces 14 fare products and non-travel alternative as shown in Table 2. A fullfactorial experiment design has a total of 4.8 million scenarios. A fully specified model has the potential of 197 parameters to be estimated (An alternate specific constant and 13 cross effects for each alternative and one generic parameter on price). The design was optimised under the D_z -error criteria (Rose and Bliemer, 2005) with all prior parameter values equal to zero. After a 24-hour optimisation run using a genetic algorithm (Olaru, Smith and Wang 2011) a 64-scenario design with a D_z -error=2.21887 was chosen for the questionnaire. The overall design was randomised and blocked into 8 subsets with 8 scenarios for each questionnaire.

3.2 Empirical Results

In this section, we reported the estimation results of a random parameter logit (RPL) model, an RPL-availability model, and an error component model (ECM). In all three models, the price effect for the major carrier was treated as a normally distributed random variable with the heterogeneity around the mean explained by two dummy coded variables. Income enters as an indicator of low income classes with monthly salary less than ¥5000. Passenger indicated the number of days that they had booked their ticket prior to departure. A prebooking variable enters the model as 1 if the passenger booked their ticket within one week of departure, and zero otherwise.

The availability model with 14 alternatives has a total of 197 parameters to be estimated. As such preliminary estimations were trialled based on some a-priori expectations of where cross-effect may be stronger. The following criteria were a guide to the modelling strategy.

- a) the cross effects between fare products on the same flights;
- b) the cross effects between the same fare classes on parallel flights; and
- c) the cross effects between different fare classes on parallel flights.

Positive cross effects based on the 'presence' of an alternative are difficult to interpret in terms of random utility maximisation. These coefficients indicate that adding an alternative to the choice set increases the probability of choice for some alternatives. In the initial investigations using a) to c) above we found five positive cross-effects significant at the 5% level. It possible for these to be reflective of some choice behaviour in that the presence of a more expensive ticket may make the difference between two cheaper alternatives less imposing. However, there was no clear pattern among the five. The positive cross-effect with the highest T-value was the presence of a non-flexible afternoon flight on the choice of premium economy at night. The final RPL-Availability model is limited to negative presence coefficients.

Table 3 listed the estimations of the price effects for the RPL and RPL-availability model, the complete estimation results are available from the corresponding author on request. The non-business customers are sensitive to price, especially to the fare products offered by the low cost carrier. Compared to higher income groups, people with low to medium income are more sensitive to price. Moreover, in both the RPL and the RPL-availability models there is weak evidence that customers who booked their current flight within one week of departure are less price sensitive than those who booked earlier. The

Table 3 –Estimation Results for RPL and RPL-Availability model

Table 4 lists the presence/absence coefficients for the RPL-Availability model significant at the 5% level. Apart from the first presence term (SFnig in LCMor) the substitutions are intuitive. The presence of a cheaper ticket at the same time of day tends to lure passengers from the more expensive ticket (i.e. LCMor effects NFMor and NFAft effects SFAft). This is known as buy down behaviour. Also the presence of similar class tickets in adjacent time slots creates a greater level of switching between flights – buy across behaviour.

Choice Οf	in the presence of	Coefficient	T-value
LCMor	SFNig	-0.432	2.86
NFMor	LCMor	-0.906	4.81
NFMD	LCMor	-0.335	2.31
NFAft	NfNig	-0.333	2.20
NFNig	NFMD	-0.632	2.04
NFNig	NFAft	-0.478	1.79
SFAft	NFaft	-0.498	1.96
SFAft	SFNig	-0.648	2.64
PEMD	PEMor	-0.688	2.34
PEAft	PEMD	-0.714	3.88

Table 4 – Availability Effects for Non-Business Travellers

The logic behind these substitutions points to correlations in the unobserved components and therefore a modelling strategy that involves nesting the alternatives. Because the data is a quasi-panel created by repeated SP choices an error components model, rather than a nested logit, is used. The error components were included to capture correlation between ticket classes and for the departure times. These are displayed in Table 5.

Table 5 – The Error Structure in the ECM

Two error component models are estimated. The first is the error component extension to the non-availability RPL model in Table 3. The second keeps the availability parameters in the estimation. Hence, we are able to see if the availability (presence) parameter estimates were picking up some of the correlation in the unobserved utilities. If this is the case, the standard errors on the availability parameters will be larger when unobserved correlation is included in the model.

The results in Table 6 show that the nesting-structure for preferences has a better fit to the data than the RPL models. However, the inclusion of availability parameters still adds some power to the estimating model.

	RPL			RPL-Availability	
	Estimates	T-test	Estimates	T-test	
Price (CNY)					
Price (Major)	$-.0175***$	10.4	$-0.0155***$	9.35	
Price (LC)	$-0.0408***$	9.59	$-0279***$	10.77	
Heterogeneity in mean (Price for the major carrier)					
BKL	.000	0.13	$.0021*$	1.79	
INC	$-.0054***$	8.91	$-0.0075***$	6.31	
Std. Devs. (Price for the major carrier)					
Price (major)	0.0074***	17.71	.0092***	15.05	
Heterogeneity in mean (Price for the major carrier)					
E01 Non-Flex	$3.42***$	18.40	$3.78***$	16.99	
E02 Semi-Flex	$1.67***$	10.61	$1.90***$	11.19	
E03 Premium	$3.12***$	9.97	$2.03***$	6.80	
E04 Morning	$2.06***$	15.63	$2.53***$	14.52	
E05 Day	$1.53***$	11.45	$1.24***$	9.84	
E06 Night	$1.36***$	8.24	$1.11***$	8.72	
Observations	2507		2507		
LL	-2953.90		-2929.85		
R^2	0.555		0.558		
χ^2 10 d.f. 5% level = 18.3			χ^2 test of for the significance of availability terms 48.10***		

Table 6 –Estimation Results for RPL and RPL-Availability model

The standard errors for all presence coefficients increased in magnitude by an average 30% (not shown here). Six of the ten parameters are no longer significant at the 5% level, shown in Table 7. However, the first listed coefficient has a larger T-value than its corresponding value in the RPL-Availability model. This parameter is anomalous because it is the only presence effect not subsumed by the error component specification.

Table 7 – Availability Effects for Non-Business Travellers

Table 8 lists the direct price elasticities for the above models. Adding availability parameters to the RPL and the Error Components Model makes little difference to the elasticities; the first two and the last two rows are fairly consistent. However, there is a substantive difference between the models of the unobserved effects.

Table 8 Direct Price Elasticity of Non-business Travellers

Availability of ticket classes has a bearing on where respondents direct their preferences. For both the RPL and the ECM estimates the inclusion of presence (cross-effect) parameters led to a better fir to the data. However, it clear that much of the specific substitution between two tickets is explained by decomposing the error into a 'fare class' component and a 'departure time' component. The error-components model was able to capture the preference structure better than that of the inclusion of presence coefficients in the RPL model. An alternate strategy – not shown here – is to specify the presence coefficients as random parameters. However, the strategy potentially leads to many random parameters to be estimated.

4 CONCLUSION

A choice modelling exercise is undertaken for short haul, non-business travel in the south east of China. Respondents answered eight replications of choice scenarios generated by an availability design. The availability design differs from fixed choice set designs in one important way: The observed correlation between (dependence of) price on quality does not have to be artificially suspended to force trade-offs. In this example the airline's ticket price hierarchy is maintained in the choice experiment. The results gave robust estimates for the price responsiveness. In addition the availability design was able to capture substitution patterns that appear to be quite reasonable. Respondents were more likely to buy down – book the cheaper of two tickets when available on the same flight or time of day or to buy across – choose the next closest time slot when their preference was not available.

BIBLIOGRAPHY

- Batsell, R. R., & Polking, J. C. (1985). A New Class of Market Share Models. Marketing Science, 4(3), 177-198.
- Belobaba, P., Odoni, A., & Barnhart, C. (Eds.). (2009). The Global Airline Industry. UK: Wiley.
- Bliemer, M. C. J., J. M., Rose and D. A., Hensher (2009a) Efficient stated choice experiments for estimating nested logit models, Transportation Research B, 43 19-35.
- Bliemer, M. C. J., J. M., Rose and S., Hesse (2009b) Approximation of Bayesian efficiency in experimental designs, Journal of Choice Modelling, 1(1) 98-127.
- Bliemer, M., & Rose, J. (2011). Experimental Design Influences on Stated Choice Outputs: An Empirical Study in Air Travel Choice. Transportation Research Part A, 45, 63-79.
- Brownstone, D., & Train, K. (1999). Forecasting New Product Penetration with Flexible Substitution Patterns. Journal of Econometrics, 89, 109-129.
- Bunch, D. S., Louviere, J. J., & Anderson, D. (1996). A Comparison of Experimental Design Strategies for Multinomial Logit Models: The Case of Generic Attributes.
- Carlsson, F. and P., Martinsson (2003) Design techniques for stated preference methods in health economics, Health Economics, 12 281-194.
- Collins, A., Rose, J., & Hess, S. (2010). Search Based Internet Surveys: Airline Stated Choice. Working Paper ITLS-WP-10-01, ITLS Sydney and ITS Leeds.
- Ferrini, S. and R., Scarpa (2007) Designs with a priori information for nonmarket valuation with choice experiments: A Monte Carlo study, Journal of Environmental Economics and Management, 53 243-363.
- Huber, J., & Zwerina, K. (1996). The Importance of Utility Balance in Efficient Choice Designs. Journal of Marketing Research, 33(3), 307-317.
- Jain, D., & Bass, F. (1989). Effect of Choice Set Size on Choice Probabilities: An Extended Logit Model. International Journal of Research in Marketing, 6, 1-11.
- Kanninen, B. J. (2002) Optimal design for multinomial choice experiments, Journal of Marketing Research, 39 214-217.
- Kessels, R., B., Jones, P., Goos and M., Vandebroek (2008) Recommendations on the use of Bayesian optimal designs for choice experiments, Quality and Reliability Engineering International, 24(6) 717-744, Special Issue: The European Network for Business and Industrial Statistics (ENBIS).
- Lazari, A. G., & Anderson, D. A. (1994). Designs for Discrete Choice Set Experiments for Estimating Both Attribute and Availability Cross Effects. Journal of Marketing Research, 31(3), 375-383.
- Louviere, J. J., Hensher, D. A., & Swait, J. D. (2000). Stated Choice Methods Analysis and Application. New York: Cambridge University Press.
- McFadden, D. (1974). Conditional Logit Analysis of Qualitative Choice Behaviour. In P. Zarembka, Frontiers in Econometrics (pp. 105-142). New York: Academic Press.
- McFadden, D. (1975). On Independence, Structure and Simultaneity in Transportation Demand Analysis. Working Paper#7511.
- Olaru, D. and Smith, B. and Wang J. (2011) "Optimal discrete choice experimental designs using genetic algorithms", paper presented at the International Choice Modelling Conference, Oulton Hall, UK, July 4-6
- Proussaloglou, K., & Koppelman, F. (1999). The Choice of Air Carrier, Flight, and Fare Class. Journal of Air Transport Management, 5, 193-201.
- Rose, J. M., & Bliemer, M. C. (2004). The Design of Stated Choice Experiments: The State of Practice and Future Challenges. working paper.
- Rose, J. M., & Bliemer, M. C. (2005). Constructing Efficient Choice Experiments. Working Paper ITLS-WP-05-07.
- Rose, J. M., & Bliemer, M. C. (2009). Constructing Efficient Stated Choice Experimental Designs. Transport Reviews, 29(5), 587-617.
- Rose, J. M., & Bliemer, M. C. (2010). Stated Choice Experimental Design Theory: the Who, the What and the Why. Working Paper.
- Ryan, M., K., Gerard and M., Amaya-Amaya (Eds.) (2007) Using Discrete Choice Experiments to Value Health and Health Care, Springer, Dordrecht, The Netherlands, ISBN: 978-1-4020-4082-5.
- Sándor, Z. and M., Wedel (2002) Profile construction in experimental choice designs for mixed logit models, Marketing Science, 21(4) 455-475.
- Timmermans, H., & Borgers, A. (1991). Mother Logit Analysis of Substitution Effects in Consumer Shopping Destination Choice. Journal of Business Research, 23, 311-323.
- Train, K. (1986). Quantitative Choice Analysis. Cambridge, MA: MIT Press.