How much airlines passengers are willing to pay in the USA: estimation of a discrete choice model

Margarita María Mejía Jaramillo

1 Introduction and objective.

This case study is based in a discrete choice experiment conducted originally by Boeing Commercial Airplanes in the United States in 2004 and reported on the paper on the Journal of Revenue and Price Management (Garrow, Jones, & Parker, 2007).

The goal with this stated preference experiment is to determine how sensitive air passengers are to various attributes of an airline itinerary, including fare, travel time, transfers, and legroom. The survey was designed with the help of Jordan Louviere of the University of Technology, Sydney.

The survey was administered to customers of an Internet airline booking service that searches for low-cost travel deals. While waiting for the search engine to return real itineraries for their requested travel, randomly selected customers were asked to complete a survey customized to their origin and destination.

Overall, the experiment aimed to gather data on consumer preferences and decision-making processes regarding air travel. The results could be used to inform marketing strategies for airlines, improve the design of flight itineraries, and provide insights into consumer behavior in the air travel industry.

2 Methodology.

2.1 Dataset

The discrete choice experiment was designed to assess the preferences of respondents regarding different flight itineraries based on various attributes such as airline, fare, level of service (non-stop, single connection on the same airline, single connection involving a change of airline), arrival and departure times, airplane type, and seat pitch (or amount of legroom).

Each respondent was shown one choice set and asked to rank the three alternatives offered based on their preferences. Additionally, the respondents were asked to indicate whether they would fly, not take the trip, or take the trip by a different mode if these were the only three air alternatives available.

Figure 1: Survey Design Example

FEATURES	Non-Stop (Option 1)	1 Stop (Option 2)	1 Stop (Option 3)
Departure time (local)	6:00 PM	4:30 PM	6:00 PM
Arrival time (local)	8:14 PM	8:44 PM	9:44 PM
Total time in air	4 hr 14 min	4 hr 44 min	4 hr 44 min
Total trip time	4 hr 14 min	6 hr 14 min	5 hr 44 min
Legroom	typical legroom	2-in more of legroom	4-in more of legroom
Airline [Airplane]	Depart Chicago Continental Airlines [B737] to San Diego	Depart Chicago Southwest Airlines [A320], connecting with Southwest Airlines [MD80] to San Diego	Depart Chicago Northwest Airlines [MD80], connecting with American Airlines [DC9] to San Diego
Fare	$565	$485	$620
1. Which is MOST attractive?	Option 1	Option 2	Option 3
2. Which is LEAST attractive?	Option 1	Option 2	Option 3

3. If these were the ONLY three options available, I would NOT make this trip by air. Yes No

4. If you would NOT take this trip by air, how else would you make this trip? Car

For this study, the highest ranked alternative is going to be treated as the choice.

This dataset was taken from the GitHub repository from Michel Bierlaire (Bierlaire, 2020). It contains 3610 observations and 56 columns. Observations with missing values were deleted, ending with a dataset of 2897 observations. The variables consider on the project are:

Table 1: Description of attributes taken into consideration for the models.

Variable	**Levels**	**Description**
SubjectId	N/A	Id
Choice	3	Indicates the chosen option. This is a label experiment 1= Direct Flight 2= 1-stop flight 3= 1-stop Flight. Change of airline.
Flight time	N/A	In hours.
Stop penalty[1]	4	0.5, 1, 1.5, 2 hours. Only alternative 2 and 3 includes are a 1-stop flight and include a stop penalty.
Difftime[2]	N/A	In hours. It was calculated as the absolute deviation between the departure time and ideal departure time or between arrival time and ideal arrival time, depending on whether arrival or departure was deemed more important.
Legroom	4	Dummy variable indicating whether the amount of legroom is: 1= 2 inches less than typical 2= typical 3= 2 inches more than typical 4= 4 inches more than typical.
Fare	N/A	In dollars.

Table 2: Description of demographics

Variable	**Levels**	**Description**
BusinessTrip	2	0= Leisure Trip 1=Business Trip
Age	4	18-24, 25-44, 45-54, + 55 years old.
Education	3	High School, Undergraduate, Postgraduate.
Income	N/A	Values are in thousand dollars
Gender	2	0=Male 1=Female
PartySize	5	Number of people traveling.
IncomeDummy[3]	2	0= Income >= 110 1=Income < 110

[1] This variable was created as the difference between the Total Flight Time – Non-Stop Flight.
[2] This variable was created from variables of the original dataset.
[3] IncomeDummy variable was created after analysis of choices to allow interactions.

2.2 Models used and utility formulation.

2.2.1 Model 1: Multinomial Logit

General Utility

$$AcsU = ASC_j + B_{StopPenaltyHours} * StopPenaltyHours + B_{Fare} * Fare + B_{Legroom2} * Legroom2 + B_{Legroom3} * Legroom3 + B_{Legroom4} * Legroom4 + B_{DiffTime} * DiffTime + \varepsilon_j$$

Specific Utilities

$$U_{NonStop} = B_{Fare} * Fare1 + B_{Legroom2} * Legroom2 + B_{Legroom3} * Legroom3 + B_{Legroom4} * Legroom4 + B_{DiffTime} * DiffTime$$

$$U_{OneStop} = ASC_{OneStop} + B_{StopPenaltyHours} * StopPenaltyHours2 + B_{Fare} * Fare2 + B_{Legroom2} * Legroom22 + B_{Legroom3} * Legroom23 + B_{Legroom4} * Legroom24 + B_{DiffTime} * DiffTime2$$

$$U_{OneStopChange} = ASC_{OneStopChange} + B_{StopPenaltyHours} * StopPenaltyHours3 + B_{Fare} * Fare3 + B_{Legroom2} * Legroom32 + B_{Legroom3} * Legroom33 + B_{Legroom4} * Legroom34 + B_{DiffTime} * DiffTime3$$

2.2.2 Model 2: Multinomial Logit with Covariates

Alternative specific constant and parameters using socio-demographics interactions found to be significant

$$ACS_{1StopChange\_Income} = ASC_{OneStopChange} + ASC_{OneStopChange\_Shift\_Income} * IncomeDummy$$

$$B_{StopPenalty\_Business} = B_{StopPenaltyHours} + B_{StopPenaltyHours\_shift\_Business} * BusinessTrip$$

Specific Utilities

$$U_{NonStop} = B_{Fare} * Fare1 + B_{Legroom2} * Legroom12 + B_{Legroom3} * Legroom13 + B_{Legroom4} * Legroom14 + B_{DiffTime} * DiffTime1$$

$$U_{OneStop} = ASC_{OneStop} + B_{StopPenalty\_Business} * StopPenaltyHours2 + B_{Fare} * Fare2 + B_{Legroom2} * Legroom22 + B_{Legroom3} * Legroom23 + B_{Legroom4} * Legroom24 + B_{DiffTime} * DiffTime2$$

$$U_{OneStopChange} = ASC_{OneStopChange\_Income} + B_{StopPenalty\_Business} * StopPenaltyHours3 + B_{Fare} * Fare3 + B_{Legroom2} * Legroom32 + B_{Legroom3} * Legroom33 + B_{Legroom4} * Legroom34 + B_{DiffTime} * DiffTime3$$

2.2.3 Model 3: Mixed Multinomial Logit

Random Utilities to allow to random taste variation in population. Both parameters are defined to follow a log negative normal distribution due to be found negative in previous models.

$$B_{StopPenaltyHours} = -\exp(\text{mu\_log\_}StopPenaltyHours + \text{sigma\_log\_}StopPenaltyHours * \text{draws\_}StopPenaltyHours)$$

$$B_{Difftime} = -\exp(\text{mu\_log\_DiffTime} + \text{sigma\_log\_DiffTime} * \text{draws\_DiffTime})$$

Specific Utilities

$$U_{NonStop} = B_{fare} * Fare1 + B_{legroom2} * Legroom12 + B_{legroom3} * Legroom13 + B_{legroom4} * Legroom14 + B_{Difftime} * DiffTime$$

$$U_{OneStop} = ASC_{OneStop} + B_{StopPenalty} * StopPenaltyHours_2 + B_{fare} * Fare2 + B_{legroom2} * Legroom22 + B_{legroom3} * Legroom23 + B_{legroom4} * Legroom24 + B_{Difftime} * DiffTime$$

$$U_{OneStopChange} = ASC_{OneStopChange} + B_1 * StopPenaltyHours + B_{fare} * Fare3 + B_{legroom2} * Legroom32 + B_{legroom3} * Legroom33 + B_{legroom4} * Legroom34 + B_{Difftime} * DiffTime$$

3 Results and discussion

3.1 Pre-estimation analysis of choices

It was run some t-test to determine whether the characteristics of individuals choosing specific alternatives differ across the three labelled alternatives.

Table 3: T-test for analysis of choices.

Variable	Statistic	NonStop	OneStop	OneStopChange
Business Trip	mean when alt is chosen:	0,3121	0,2484	0,2176
	mean when alt is not chosen:	0,2337	0,2952	0,2998
	t-test (mean if chosen - mean if not chosen)	4,48	-2,14	-3,75
PartySize	mean when alt is chosen:	1,6819	1,7263	1,6898
	mean when alt is not chosen:	1,7089	1,6833	1,6905
	t-test (mean if chosen - mean if not chosen)	-0,68	0,85	-0,01
Income	mean when alt is chosen:	111,5025	107,1053	91,3079
	mean when alt is not chosen:	99,581	107,9005	110,6552
	t-test (mean if chosen - mean if not chosen)	3,56	-0,18	-5,08
Female	mean when alt is chosen:	0,4628	0,4358	0,4375
	mean when alt is not chosen:	0,4366	0,4583	0,4576
	t-test (mean if chosen - mean if not chosen)	1,32	-0,9	-0,78
Undergraduate	mean when alt is chosen:	0,4804	0,4842	0,4861
	mean when alt is not chosen:	0,4851	0,4814	0,4811
	t-test (mean if chosen - mean if not chosen)	-0,24	0,11	0,19
Posgraduate	mean when alt is chosen:	0,3362	0,3411	0,3704
	mean when alt is not chosen:	0,355	0,3423	0,3371
	t-test (mean if chosen - mean if not chosen)	-0,99	-0,05	1,32
Age25-44	mean when alt is chosen:	0,609	0,5853	0,6319
	mean when alt is not chosen:	0,6075	0,6131	0,6045
	t-test (mean if chosen - mean if not chosen)	0,08	-1,13	1,09
Age45-54	mean when alt is chosen:	0,201	0,1811	0,1435
	mean when alt is not chosen:	0,1632	0,1908	0,1972
	t-test (mean if chosen - mean if not chosen)	2,49	-0,5	-2,87
Age+55	mean when alt is chosen:	0,0985	0,1221	0,0972
	mean when alt is not chosen:	0,1103	0,0983	0,103
	t-test (mean if chosen - mean if not chosen)	-0,95	1,47	-0,37

The output shows that Variable BusinessTrip is statistically significant for respondents who choose alternative 1 (direct flight); Respondents with an average income $111.502 dollars are more likely to choose alternative 1, while those with an average Income $110.655 tends not to go with alternative 3 (1-stop flight with airline change); Group Age 45-54 seems to be more likely to go for alternative 1, and not to choose alternative 3. All other variables are not statistically significant with a confidence 95%

3.2 Model Results

3.2.1 Results from Model 1: Multinomial Logit

All parameters specified in the utility function for the model are significant. Results are reported in table 4.

The alternative 1 (asc_NonStop = 0) is set as the baseline. This means that comparing if someone chooses alternative 2, there is a decrease in the utility of -1,36, otherwise if alternative 3 is chosen, utility goes down by -1,57.

All parameters have a negative value, except for Legroom, that when you increase the space compared to the baseline (2 inches less than typical), the utility increases.

Table 4: Parameters estimated for model 1Multinomial Logit

	Estimate	**Std.err.**	**t-ratio(0)**	**p(2-sided)**	**Rob.std.err.**	**Rob.t-ratio(0)**	**Rob.p-val(0)**
asc_Non Stop	0	NA	NA	NA	NA	NA	NA
asc_1Stop	-1,361345	0,10743197	-12,671694	0	0,10602998	-12,839246	0
asc_1Stop Change	-1,5657577	0,10880917	-14,389942	0	0,10688444	-14,649071	0
b_StopPenaltyHours	-0,3411405	0,0739853	-4,610922	4,01E-06	0,07422614	-4,5959612	4,31E-06
b_Fare	-0,0193137	0,00076294	-25,314747	0	0,00088533	-21,815218	0
b_Legroom 2	0,37965766	0,08730273	4,34874904	1,37E-05	0,08931791	4,25063285	2,13E-05
b_Legroom 3	0,42743215	0,08611748	4,96336132	6,93E-07	0,0899616	4,7512734	2,02E-06
b_Legroom 4	0,65897922	0,08756869	7,52528315	5,26E-14	0,09280937	7,10035248	1,24E-12
b_DiffTime	-0,1082203	0,01139836	-9,4943755	0	0,01209442	-8,9479523	0

People is willing to pay $ 19,66 dollars for a Legroom 2 typical, $22.13 for a Legroom3 (2 inches more than the typical), and $34.12 for a Legroom 4 (4 inches more than the typical). They are also willing to pay $5,60 dollars for each decrease of an hour to their desire departure/arrival time.

3.2.2 Results from Model 2: Multinomial Logit with covariates

All parameters specified in the utility function for the model are significant. Results are reported in table 5.

Table 5: Parameters estimated for model 2: multinomial logit with covariates

	Estimate	Std.err.	t-ratio(0)	p(2-sided)	Rob.std.err.	Rob.t-ratio(0)	Rob.p-val(0)
asc_NonStop	0	NA	NA	NA	NA	NA	NA
asc_1Stop	-1,3601509	0,10792586	-12,602641	0	0,10637183	-12,786759	0
asc_1StopChange	-1,8587099	0,13759878	-13,508185	0	0,13810748	-13,45843	0
asc_1StopChange Shift_Income	0,47003375	0,1249758	3,76099812	0,00016924	0,12702485	3,70032922	0,00021532
b_StopPenaltyHours	-0,2653509	0,07613242	-3,4853869	0,00049143	0,07596118	-3,4932437	0,00047719
b_StopPenaltyHours_Shift_Business	-0,3129887	0,08016643	-3,9042371	9,45E-05	0,08337898	-3,7538089	0,00017417
b_Fare	-0,0195091	0,0007699	-25,339698	0	0,00089618	-21,769257	0
b_Legroom2	0,3844103	0,08774192	4,38114767	1,18E-05	0,08976771	4,28227824	1,85E-05
b_Legroom3	0,42430099	0,08639439	4,91120989	9,05E-07	0,08974497	4,72785262	2,27E-06
b_Legroom4	0,66959854	0,08805202	7,60457894	2,86E-14	0,09297238	7,20212356	5,93E-13
b_DiffTime	-0,1091328	0,01147551	-9,510054	0	0,01221468	-8,934555	0

3.2.3 Model Comparison

Model 1 and 2 were compared through a Likelihood Ratio Test. Results are reported in table 6. Null Hypothesis is rejected, which means that model 2 is statistically significantly better than model 1.

Table 6: Likelihood Ratio Test

	LL	Parameters
Model1: MNL	-1894,51	8
Model2: MNL_Covariates	-1878,04	10
Difference	16,47	2
Likelihood ratio test-value:		32,94
Degrees of freedom:		2
Likelihood ratio test p-value:		7,03E-08

3.2.4 Elasticity for model 2

Values are Reported on Table 7. When the price for Direct Flight increases by 1%, probability of being chosen decreases by 1,874 %, while alternative 2 goes down by 3,957 % and the alternative 3 gains 4,035 %.

When the price for 1-Stop Flight increases by 1%, probability of being chosen decreases by 4,868 %, while alternative 1 goes up by 0,868 and the alternative 3 gains 1,202.

When the price for 1-Stop Flight Change increases by 1%, probability of being chosen decreases by 4,999 %, while alternative 1 goes up by 0,797 % and the alternative 2 gains 1,079 %.

Table 7: Elasticity prediction for model 2.

	Direct Flight	**1Stop Flight**	**1Stop Flight Change**
Elasticity 1% ↑ Fare Direct Flight	-1,874	3,957	4,035
Elasticity 1% ↑ Fare 1Stop Flight	0,868	-4,868	1,202
Elasticity 1% ↑ Fare 1Stop Flight Change	0,797	1,079	-4,999

3.2.5 Results Model 3: Mixed Multinomial Logit

Results from Mixed Multinomial Logit model are reported in table 8.

All parameters specified in the utility function for the model are significant except for the mu_log_StopPenaltyHours. On average, the variable StopPenaltyHours has not impact on model (=0), but for some people matters (because sigma_log_StopPenaltyHours is significant). On the other hand, mu_log_DiffTime and sigma_log_DiffTime are significant, which means there is heterogeneity in the population. The average impact of variable Difftime is

–0,5659672.

Table 8: Parameters estimated for model 3: Mixed Multinomial Logit

	Estimate	**Std.err.**	**t-ratio(0)**	**p(2-sided)**	**Rob.std.err.**	**Rob.t-ratio(0)**	**Rob.p-val(0)**
asc_NonStop	0	NA	NA	NA	NA	NA	NA
asc_1Stop	-0,9634695	0,18155113	-5,306877	1,12E-07	0,1837542	-5,2432517	1,58E-07
asc_1StopChange	-1,2571074	0,18099145	-6,9456728	3,77E-12	0,1845821	-6,8105594	9,72E-12
mu_log_StopPenaltyHours	-0,6037843	0,48469933	-1,2456883	0,21287884	0,50349499	-1,1991863	0,23045551
sigma_log_StopPenaltyHours	-3,7565434	0,6725754	-5,585312	2,33E-08	0,70006937	-5,3659589	8,05E-08
b_Fare	-0,0357908	0,0023283	-15,372082	0	0,00299785	-11,938808	0
b_Legroom2	0,5281755	0,13495033	3,91385117	9,08E-05	0,13377116	3,94835092	7,87E-05
b_Legroom3	0,61701137	0,139932	4,4093657	1,04E-05	0,14351367	4,29932115	1,71E-05
b_Legroom4	0,96558833	0,13561669	7,11998134	1,08E-12	0,14207961	6,79610744	1,07E-11
mu_log_DiffTime	-3,0710841	0,36707577	-8,3663494	0	0,37692208	-8,1477957	4,44E-16

sigma_log_DiffTime	-2,2368974	0,31332427	-7,1392409	9,39E-13	0,32311741	-6,9228627	4,43E-12

4 References

Bierlaire, M. (2020, December 27). Retrieved from Michel Bierlaire's Github Repository: https://github.com/michelbierlaire/biogeme/blob/master/examples/assisted/airline.dat

Garrow, L. A., Jones, S. P., & Parker, R. A. (2007). How much airlines customers are willing to pay: an analysis of price sensitivity in online distribution channels. *Journal of Revenue and Pricing Management, 5*, 271-290.