# Itinerary Choice Model

Because itinerary choice modelling is important in estimating how passengers will respond to aviation policy that may involve different cost changes on different itineraries, this supplementary information gives more details on how itinerary choice is modelled in the main paper, the parameters used, and the model estimation process.

Itinerary choice models in aviation seek to express how passengers choose between different routing options. For example, travel between a given city pair may involve a direct flight or connecting flights via a hub airport; for many cities, passengers must also choose which airport they will travel from, to or via. Historically, this problem has been studied from various different perspectives, examining factors such as the day of the week, time of flight or level of customer service on passenger choices (e.g. Coldren at al., 2003; Coldren & Koppelman, 2005; Koppelman et al., 2008; Adler et al., 2005; Warburg et al., 2006; Theis et al., 2006). However, this disaggregate approach often results in models with low predictive power, with typical R2 values of 0.2-0.5. For the purposes of the AIM modelling system and this paper, a simpler model which provides estimates based on the relative cost, journey time, frequency and other aggregate journey attributes is more appropriate. We also use lagged airport-level demand to capture the impact of airport characteristics which affect passenger airport choice. This enables the use of the model for future projections and to assess hypothetical scenarios where limited information may be available.

As illustrated in the main paper, we use a multinomial logit formulation to estimate yearly aggregate itinerary choice. Logit models are widely used to model choices between different alternatives, both in an aviation context and more generally (e.g. Ben-Akiva, 1985; Proussaloglou & Koppelman, 1999; Coldren et al. 2003). An itinerary is considered to be a series of connected flight legs between a given origin and destination city. Two itineraries are considered to be distinct if they have different origin, destination or connecting airports. Table S1 gives an example of itineraries available in 2015 for the city-pair New York to Istanbul and their characteristics. In our model, the number of passengers *Dijk* between cities *i* and *j* on itinerary *k* is represented as

 ,

where the deterministic part of the utility, *Vijk*, for an itinerary *k* travelling between airport *m* in *i* and airport *n* in *j*, is:

 ,

and *fijk* is the itinerary fare, *tijk* is the total itinerary travel time, *freqijk* is the itinerary frequency, *Nlegsijk* is the number of flight legs in the itinerary, is the total number of non-transfer scheduled passengers using airport *m* in the previous year, and the parametersare estimated. Itinerary and airport-level passenger numbers and fare are derived from Sabre (2017) passenger data. This dataset includes actual routing and fare information on all global itineraries booked through the global Computer Reservation System (CRS) in 2015, as well as estimated data on all itineraries booked through alternative systems (for example, low-cost carrier flights). Itinerary frequency is defined as the minimum yearly flight frequency across all flight legs; this and itinerary journey time are derived from Sabre (2017) schedule data. A logarithmic form is used for the frequency variable to account for diminishing returns from the gain in service attractiveness for adding additional flights (Douglas & Miller, 1974). The logarithmic form can also be derived by considering each individual flight as a nested alternative with equal utility. When these alternatives are aggregated, it produces a logsum term which is proportional to the logarithm of the flight frequency (Ben-Akiva and Lerman, 1985).

Table S2 shows the model parameters by key region-pair. Using this formulation, we are able to obtain typically higher R2 values than most literature models (in the range 0.6-0.9, depending on region-pair). Standard errors are given in brackets. Nearly all parameters are significant for all regions. The parameters also behave as expected: for example, lower-cost itineraries are typically favoured over higher-cost ones; shorter journey times, higher flight frequencies and larger airports are preferred; and an itinerary with fewer flight legs is strongly preferred to one with more. Additionally, by dividing the fare and time parameters we are able to derive effective air passenger values of time which are comparable with other estimates. For example, we obtain a value of time of $42/hour in year 2015 US dollars for flights within the North American region, which can be compared to Adler et al. (2005; $38 when expressed in year 2015 dollars). Examples of projected itinerary share for a sample city-pair are also included in Table S1.

# References

Adler, T., Falzarano, C. & Spitz, G. (2005). Modeling service trade-offs in itinerary choices. *Transportation Research Record, 1915*, 20-26.

Ben-Akiva, M. E., 1985. Discrete choice analysis. MIT Press, Boston.

Coldren, G., Koppelman, F., Kasturirangan, K. & Mukherjee, A. (2003). Modeling aggregate air-travel itinerary shares: logit model development at a major (US) airline. *Journal of Air Transport Management, 9*, 361-369.

Coldren, G. & Koppelman, F. (2005). Modeling the competition among air-travel itinerary shares: GEV model development. *Transportation Research Part A, 39*, 345-365.

Douglas, G. & Miller, J., (1974). Economic Regulation of domestic air transport: theory and practice. Brookings Institution, Washington DC.

Dray L., Krammer P., Doyme K., Wang B., Al Zayat K, O’Sullivan A. & Schäfer A. (2017). AIM2015: Validation and Initial Results from an Open-Source Aviation Systems Model. 21st ATRS World Conference, July 5-8.

Koppelman, F., Coldren, G. & Parker, R. (2008). Schedule delay impacts on air-travel itinerary demand. *Transportation Research Part B,* 42, 263-273.

Prossalougou, K., & Koppelman, F. (1999). The choice of air carrier, flight and fare class. *Journal of Air Transport Management, 5*, 193-201.

Theis, G., Adler, T., Clarke, J.-P. & Ben-Akiva, M. (2006). Risk aversion to short connections in airline itinerary choice. *Transportation Research Record, 1951*, 28-36.

Warburg, V., Bhat, C & Adler, T. (2006). Modeling demographic and unobserved heterogeneity in air passenger service attributes in itinerary choice. *Transportation Research Record, 1951*, 7-16

Sabre, 2017. Market Intelligence database. <https://www.sabreairlinesolutions.com/home/software_solutions/product/market_competitive_intelligence/>

Table S1. Example itineraries, New York to Istanbul

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Origin Airport | Destination Airport | Hub Airport | Journey time, min | One-way fare, year 2015 USD | Minimum leg frequency, flights/year | Modelled passengers, 2015 |
| JFK | IST | - | 590 | 740 | 930 | 59000 |
| JFK | IST | CDG | 710 | 520 | 2480 | 8500 |
| *JFK* | *IST* | *LHR* | *710* | *780* | *2800* | *8600* |
| JFK | IST | FCO | 720 | 420 | 1100 | 6500 |
| JFK | IST | SVO | 840 | 360 | 810 | 4600 |
| JFK | IST | FRA | 700 | 760 | 1400 | 7100 |
| JFK | IST | AMS | 710 | 540 | 1100 | 6600 |
| JFK | IST | ZRH | 700 | 520 | 960 | 6400 |
| JFK | IST | KBP | 770 | 340 | 110 | 2800 |

Table S2. Key itinerary choice model parameters

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Route group | Intercept | Fare  | Time  | Frequency | Nlegs | Paxorigin | Paxdest | R2 |
| Intra North America | 0.86 (0.004) | -3.9e-03 (6e-05) | -5.5e-03 (5e-05) | 0.74 (0.004) | -1.99 (0.01) | 2.8e-08 (4e-10) | 2.8e-08 (4e-10) | 0.59 |
| Intra Europe | 0.76 (0.006) | -5.1e-03 (9e-05) | -2.8e-03 (4e-05) | 0.84 (0.004) | -3.43 (0.02) | 3.9e-08 (9e-10) | 4.0e-08 (9e-10) | 0.65 |
| Intra Asia | 0.95 (0.01) | -2.1e-03 (1e-04) | -1.3e-03 (5e-05) | 0.82 (0.009) | -3.51 (0.02) | 3.5e-08 (1e-09) | 3.6e-08 (1e-09) | 0.58 |
| Intra South America | 0.81 (0.02) | -8.2e-03 (4e-04) | -1.5e-03 (2e-04) | 0.88 (0.02) | -2.50 (0.06) | 1.2e-07 (7e-09) | 1.1e-07 (7e-09) | 0.60 |
| Intra Middle East | 0.67 (0.04) | -3.4e-03 (5e-04) | -2.7e-03 (2e-04) | 0.60 (0.03) | -2.88 (0.08) | 5.2e-08 (5e-09) | 5.2e-08 (5e-09) | 0.91 |
| Intra Africa | 0.97 (0.03) | -1.6e-03 (2e-04) | -3.6e-04 (7e-05) | 0.53 (0.03) | -1.30 (0.07) | 3.9e-07 (4e-08) | 3.5e-07 (4e-08) | 0.85 |
| N. America - Europe | 0.84 (0.006) | -8.5e-04 (4e-05) | -3.7e-03 (5e-05) | 0.72 (0.006) | -2.24 (0.02) | 8.2e-08 (1e-09) | 8.1e-08 (1e-09) | 0.91 |
| N. America - Asia | 1.13 (0.01) | 2.7e-04 (5e-05) | -2.6e-03 (6e-05) | 0.78 (0.02) | -2.47 (0.03) | 2.4e-08 (1e-09) | 2.2e-08 (1e-09) | 0.80 |
| Europe - Asia | 1.00 (0.01) | -5.7e-04 (4e-05) | -2.3e-03 (3e-05) | 0.79 (0.009) | -2.36 (0.02) | 1.9e-08 (9e-10) | 2.1e-08 (1e-09) | 0.90 |
| Europe – M. East | 0.94 (0.01) | -2.2e-03 (8e-05) | -2.5e-03 (7e-05) | 0.72 (0.09) | -2.58 (0.03) | 5.6e-08 (2e-09) | 5.6e-08 (2e-09) | 0.91 |
| Asia - Middle East | 0.91 (0.15) | -1.8e-04 (1e-04) | -2.6e-03 (7e-05) | 0.69 (0.17) | -1.91 (0.03) | 3.2e-08 (2e-09) | 3.3e-08 (2e-09) | 0.93 |
| Africa - Europe | 1.01 (0.01) | -7.9e-04 (5e-05) | -1.6e-03 (5e-05) | 0.66 (0.01) | -2.27 (0.03) | 6.3e-08 (2e-09) | 6.6e-08 (3e-09) | 0.74 |