www.pwc.co.uk

International Competition: Hubs modelling

Airports Commission

Final report

January 2014

Contents

Important Notice

This final document has been prepared for the Airports Commission in accordance with the terms of the Airports Commission Analysis and Strategy Support framework and the Contract Reference PPRO 04/08/72 dated 2nd May 2013 and solely for the purpose and on the terms agreed with the Airports Commission within the Project Inception Document reference D3. We accept no liability (including for negligence) to anyone else in connection with this document.

This document contains information obtained or derived from a variety of third party sources as indicated within the document. PwC has not sought to establish the reliability of those sources or verified the information so provided.

Should any person other than the Airports Commission obtain access to and read this document, such person accepts and agrees to the following terms:

1. The reader of this document understands that the work performed by PwC was performed in accordance with instructions provided by our client, the Airports Commission, and was performed exclusively for their benefit and use. The document may therefore not include all matters relevant to the reader.

2. The reader agrees that PwC accepts no liability (including for negligence) to them in connection with this document.

Executive summary

The DfT's aviation model and the international transfer passenger market

The Department for Transport's (DfT) aviation modelling suite consists of the National Air Passenger Demand Model (NAPDM) and the National Air Passenger Allocation Model (NAPAM). The NAPDM produces demand forecasts of passengers who are arriving at, departing from, and transiting through the UK. The NAPAM allocates this national level demand across the largest UK airports based on a detailed geographic breakdown of passenger origins, surface access transport costs and service frequency modelling.

The models can be run to provide passenger demand forecasts which are either 'unconstrained' or 'constrained' by airport capacity. The former provides insight into the total potential demand for UK airports, whereas the latter provides a more realistic forecast of actual demand. By comparing the two projections, it is possible to derive an implied number of routes that are being lost due to capacity constraints.

However, the Airports Commission's (AC) Discussion Paper 01: *Aviation Demand Forecasting* highlighted that a limitation of the existing DfT aviation modelling suite was that "it does not fully capture the international transfer passenger market",[1] where this is defined to include international transfer passengers connecting via:

- a UK hub, i.e. passengers whose journeys originate and terminate outside the UK and who connect via a UK hub airport; and
- an overseas hub, i.e. passengers whose journeys originate and terminate outside the UK and who connect via a non-UK hub airport but could potentially transfer via a UK hub if an appropriate and attractively-priced service was available.

This means that the model as currently constructed only partially assesses the impact of capacity constraints on passenger demand, in particular from international transfer passengers.

Aim of the project

The AC asked PwC to construct two logit models estimated on historical passenger demand data using an approach that replicates the existing DfT approach for modelling passenger choice decisions for international transfer passengers travelling through the UK (Heathrow, Gatwick and Manchester), Paris Charles de Gaulle (CDG), Frankfurt (FRA), Amsterdam Schiphol (AMS) and Dubai (DBX). The AC asked for two models in order that we could explore the impact of including fares as an explanatory variable (i.e. the AC asked for one model to include fares and the other to exclude fares).

A summary of our conclusions

1) The international passenger transfer (logit) model should include a variable for fares

We have tested several specifications of the logit model which include some or all of the following variables:

- the value of time spent in the air;
- the value of time spent waiting for the hub connection (this is also referred to as the 'empirical frequency' variable); and

[1] https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/73143/aviation-demand-forecasting.pdf

- the fare paid by passengers for the entire route.[2]

Intuitively we would expect all three variables to be important in determining passenger choice as they each capture a different aspect of the 'cost' to the passenger. This is confirmed by our logit analysis which shows that the three coefficients are significantly different from zero, and that they all negatively influence choice (i.e. the more time that is spent in the air or waiting for the hub connection and the higher the fare, the less attractive is a particular route).

2) The fares variable should be taken from the 'Mixed' fares dataset

The AC and DfT supplied three different fares datasets to be tested in the analysis:

i. 'Synthetic' fare data – this dataset has been developed by the DfT based on modelled fares data.

ii. 'Observed' fare data – this is a dataset that contains actual fares data for routes that were chosen by passengers.

iii. 'Mixed' fare data – this is a dataset that contains observed fares where they existed and used synthetic fares for routes which were not chosen by passengers.

Our analysis suggests that the 'Synthetic' fares data performs badly in the logit models. More specifically, the fares variable enters the equation with the incorrect (i.e. positive) sign or fails to be significant, whereas the fares variable using 'Observed' and 'Mixed' fare data does the opposite (i.e. enters the equation with the correct, positive, sign and is significant in all cases).

In principle, this means that either 'Observed' or 'Mixed' datasets could be used. On balance, and notwithstanding the shortcomings of the Synthetic fare data referred to above, our preference is to use the mixed fare dataset because the logit model that uses this data provides a better fit (as measured by the "pseudo R-squared" statistic);[3] and the dataset should be richer by virtue of how it includes fare observations that were not chosen. However, we note that the choice of dataset is a matter of judgment and should also be informed by other factors such as an assessment of the underlying robustness of the data.

3) The international passenger transfer model should be estimated on the latest available data

Our analysis shows that the set of coefficients estimated on 2008 data is statistically different to the set of coefficients estimated on 2011 data. More specifically, passengers appear to be more sensitive to time spent in the air in 2011 than they were in 2008 (i.e. passengers are more averse to long flights in 2011 than they were in 2008). This same result is obtained regardless of whether we use the CAA or PaxIS datasets (which we refer to in more detail below). However, there does not appear to be such a clear cut change in passenger sensitivity to the empirical frequency (where we obtain different results in the two datasets).

This suggests that the relationships between route attributes and their impact on route choice have changed over this time period. In the absence of any clear understanding of why these changes have occurred and what changes might take place over the timeframe being considered by the Airports Commission, our inclination would be to base the model on the most contemporaneous data that is available (i.e. the 2011 data).

---

[2] The fares data used in this exercise refers to the headline ticket prices and includes mandatory costs such as airport charges and taxes but not discretionary costs such as luggage, seating, credit card charges etc.

[3] The R-squared statistic is commonly used to measure the goodness of fit in a simple ordinary-least squares (OLS) regression model. As logit models are non-linear, R-squared statistics cannot be used. To overcome this, the 'Pseudo R-squared' can be used to assess the goodness of fit of a nonlinear model. However, care needs to be taken in how one uses the Pseudo R-squared statistic. Chapter 3 of Kenneth Train's 'Discrete Choice Methods with Simulation' states that "*In comparing two models estimated on the same data and with the same set of alternatives (such that LL(0) is the same for both models), it is usually valid to say that the model with the higher ρ^2 fits the data better. But this is saying no more than that increasing the value of the log-likelihood function is preferable. Two models estimated on samples that are not identical or with a different set of alternatives for any sampled decision maker cannot be compared via their likelihood ratio index.*"

4) The AC should undertake due diligence of the CAA and PaxIS datasets to inform a final assessment of its preferred data source

As requested by the AC, we also tested two different sources for Heathrow passenger demand (i.e. the data points for flight capacity, frequency and fares for the routes in which Heathrow was chosen as an interchange hub) within the dataset:

- PaxIS (Passenger Intelligence Services) which has been developed by IATA Business Intelligence; and
- The CAA dataset which consists of passenger interview surveys, conducted and initially processed by the CAA and further processed by the DfT for use in aviation modelling.

Our analysis shows that the coefficients estimated using the two datasets are statistically different to each other and therefore that it may be important to make a choice of one over the other. However, we have no *a priori* or statistical reason to suggest that one dataset performs better than the other. Therefore, we recommend that the AC undertake a due diligence of the two datasets, taking into account factors such as the integrity of the underlying data, to inform a final assessment of its preferred data source.

Modelling methodology and recommendations

Scope and methodology

Scope

The AC asked PwC to construct two logit models estimated on historical passenger demand data using an approach that replicates the existing DfT approach for modelling passenger choice decisions for international transfer passengers travelling through the UK (Heathrow, Gatwick and Manchester), Paris Charles de Gaulle (CDG), Frankfurt (FRA), Amsterdam Schiphol (AMS) and Dubai (DBX). The AC asked for two models to explore the impact of including fares as an explanatory variable (i.e. the AC asked for one model to include fares and the other to exclude fares).

Replicating the DfT model

The AC wanted to ensure consistency with the DfT's modelling methodology as documented in the technical paper *Logit re-estimation overview* (DoNFLSB/10/014 v1.4, July 2011). We began by replicating this model, which took the form of a conditional logit model as follows:

$$U = \beta_1(VoT * Airtime) + \beta_2\left(VoT * \left((1-(1-a)^{F_2}) * \frac{8}{F_2}\right) + VoT * \left((1-(1-a)^{F_3}) * \frac{8}{F_3}\right)\right) + \varepsilon$$

Where

U is the utility of an individual passenger;

VoT is the value of time of the passenger;

Airtime is the time of air travel, measured by adding the individual airtimes of origin to hub and hub to destination;

F_2 and F_3 are the flight frequency of the first leg and the second leg of the journey respectively;[4]

a is a constant that captures passenger preference for time spent at the airport. The constant can vary by passenger type (e.g. business, leisure…) but in this model we have only one passenger type, the international transfer passenger; and

ε is the error term.

The term $\left(VoT * \left((1-(1-a)^{F_2}) * \frac{8}{F_2}\right) + VoT * \left((1-(1-a)^{F_3}) * \frac{8}{F_3}\right)\right)$ is also referred to as the *empirical frequency* in the DfT's technical paper and can be interpreted as a passengers value of time spent waiting for the hub connection.

[4] Note: we are following the DfT notation here. F_1 refers to the joint frequency of the route ($F_2 + F_3$), but is not used in our estimation of the logit models.

The coefficients of a conditional logit model should be interpreted as the marginal impact of a given variable on passenger utility. As these variables are in the form of costs, they detract from a passengers' utility. The option with the highest utility after accounting for these costs is the option that is most likely to be chosen.

The results from DfT's original model (as specified in the technical paper) and PwC's replication are shown in Table 1 below.

Table 1: Comparison between DfT and PwC's logit model

Model	DfT model	PwC model
Dependent variable	Hub choice	Hub choice
Value of airtime	-0.0402 (0.0248) z = -1.62	-0.0402 (0.0248) z = -1.62
Empirical frequency	-0.0927*** (0.00530) z = -17.49	-0.0927*** (0.00530) z = -17.49
Observations	1,769	5,129
Pseudo R-sq	0.6939	0.6939

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The results confirm the replication: the model coefficients, standard errors (in parentheses) and model fit (specified by the Pseudo R-squared) of the DfT model and the PwC model are the same. The result was confirmed in two statistical software packages: STATA and Biogeme.[5] We note that fares data was excluded from the DfT's original model due to a lack of source data.[6] We re-consider the inclusion of a fare variable using alternative datasets later in this report.

We also note that there is a difference between the number of observations reported in the original DfT model (1,769) and the PwC model (5,129). The DfT used Biogeme to estimate the model; at the request of the AC, we re-estimated the model using STATA. The difference in the number of observations reported is due to the way in which these two statistical packages define an "observation".[7] However, the models are estimated using the same dataset.

New data and fares data

We were provided with two datasets by the DfT - one from 2008 and the other from 2011 - for the seven interchange hubs. The DfT also provided us with two extra datasets in which the data for Heathrow had been replaced with PaxIS Heathrow data. Frequency information was sought from OAG and Innovata electronic timetable data held by the DfT.

We hence re-estimated the logit model using the following four datasets:

[5] The AC asked us to develop the model using STATA. However, the DfT used Biogeme to estimate their original Passenger Airport Choice Model (as documented in their technical paper 'Logit re-estimation overview'. DONFLSB/10/014 v1.4, July, 2011). Biogeme is an open source freeware designed for the estimation of discrete choice models that is widely used by many transport economists.
[6] Please refer to paragraph 8.3 onwards in the technical paper 'Logit re-estimation overview'. DONFLSB/10/014 v1.4, July, 2011.
[7] Reconciling STATA's method of weighting the observations with Biogeme's method is outside the scope of this report.

- 2008, sourced from CAA (UK) and PaxIS (overseas);
- 2011, sourced from CAA (UK) and PaxIS (overseas);
- 2008, sourced from CAA (UK excluding Heathrow) and PaxIS (overseas and Heathrow); and
- 2011, sourced from CAA (UK excluding Heathrow) and PaxIS (overseas and Heathrow).

We also tested a different specification of the model, where the fares variable is included. The AC and DfT supplied three different fares datasets to be tested in the analysis:

i. 'Synthetic' fare data for 2008 and 2011 – this dataset has been developed by the DfT based on modelled fares data. This data is useful because not all fares are captured in the 'observed' dataset, which is a particular problem when trying to capture observations for options that were available to passengers but were not chosen;

ii. 'Observed' fare data for 2011 – this is a dataset that contains actual fares data for routes that were chosen by passengers. However, the dataset only holds fare data for the hubs that are actually chosen and not for the options that were not chosen; and

iii. 'Mixed' fare data for 2011 – this is a dataset that contains observed fares where they existed and synthetic fares for routes which were not chosen by passengers.

The model outputs are presented in section 2.2.

Model outputs and statistical tests

Table 2 shows our model results using the alternative datasets but *without* the fares variable. The value of airtime and empirical frequency coefficients are broadly similar across different years and across different datasets, in both signs and magnitude. The coefficients are all significant at the 1% confidence level.

The results also indicate that passengers are more sensitive to the value of airtime in 2011 than they were in 2008 (i.e. passengers are more averse to long flights in 2011 when compared with 2008). A similar result is obtained when using either the CAA or PaxIS datasets. Later in this report, we formally test whether this difference in coefficients can be considered statistically significant.

However, passenger sensitivity to the empirical frequency variable does not appear to exhibit such a clear cut change. Passengers appear to have become slightly more averse to spending time at the interchange hub in 2011 compared with 2008 when using the CAA dataset, but the reverse is true when using the PaxIS dataset.

Table 2: Logit model estimated without fares, using different datasets

Dataset	2008, CAA	2011, CAA	2008, PaxIS Heathrow	2011, PaxIS Heathrow
Dependent variable	Hub choice	Hub choice	Hub choice	Hub choice
Value of airtime	-0.00831*** (0.00142) z = -5.86	-0.0122*** (0.00201) z = -6.09	-0.00824*** (0.00142) z = -5.79	-0.0136*** (0.00207) z = -6.57
Empirical frequency	-0.0530*** (0.00212) z = 24.98	-0.0657*** (0.00334) z= -19.66	-0.0494*** (0.00190) z = -26.04	-0.0434*** (0.00226) z = -19.17
Observations	44,148	40,627	47,212	47,600

Pseudo R-sq	0.1839	0.2081	0.1767	0.1644

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3 shows the results after estimating the model but including a fares variable using with 'Synthetic' data. We see that the value of airtime and the empirical frequency variables behave in a similar way to before but the direction, magnitude and significance of the fares variable vary across the four datasets. We interpret these results in two ways. First, the value of airtime and empirical frequency variables are important to the specification of the model. Second, if we are to include a fares variable, then it should not be done using solely 'Synthetic' data.

Table 3: Logit model estimated with 'Synthetic' fares, using different datasets

Dataset	2008, CAA	2011, CAA	2008, PaxIS Heathrow	2011, PaxIS Heathrow
Dependent variable	Hub choice	Hub choice	Hub choice	Hub choice
Value of airtime	-0.0110*** (0.00163) z = -6.76	-0.0120*** (0.00213) z = -5.62	-0.00827*** (0.00159) z = -5.20	-0.0154*** (0.00215) z = -7.18
Empirical frequency	-0.0497*** (0.00231) z = -21.52	-0.0662*** (0.00434) z = -15.24	-0.0494*** (0.00220) z = -22.47	-0.0410*** (0.00275) z = -14.94
'Synthetic' fare	0.000844*** (0.000250) z = 3.38	-9.06e-05 (0.000321) z = -0.28	9.09e-06 (0.000222) z = 0.04	0.000584** (0.000241) z = 2.42
Observations	44,148	40,627	47,212	47,600
Pseudo R-sq	0.1854	0.2081	0.1767	0.1652

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4 and Table 5 show the results of including a fares variable based 'Observed' and 'Mixed' fares data for 2011.[8] The fares coefficient has the sign we would expect (i.e. negative) and is significant. We note from the results that:

- the 'Observed' fares variable is significant at the 1% level in the PaxIS dataset but only at the 5% level in the CAA dataset, whereas the 'Mixed' fares variable is significant at the 1% level with both datasets;
- the fit of the model using 'Mixed' fares data is higher, probably due to the larger number of observations with the inclusion of options that were available, but not chosen by the passenger; and
- the fares variable coefficient is higher when using 'Mixed' data than when using 'Observed' data.

Table 4: Logit model with 'Observed' fares, estimated using different datasets

[8] Fares data was excluded from the DfT's previous analysis, which used 2008 data due to a lack of source data. The DfT and AC therefore wanted us to focus our fares analysis on 2011 data - although, for comparison with DfT's previous analysis, we did use 2008 data when analysing the potential use of 'Synthetic' data.

Dataset Dependent variable	2011, CAA Hub choice	2011, PaxIS Heathrow Hub choice
Value of airtime	-0.0127*** (0.00178) z = -7.11	-0.0130*** (0.00174) z = -7.49
Empirical frequency	-0.0466*** (0.00444) z = -10.49	-0.0403*** (0.00432) z = -9.33
'Observed' fare	-0.000227** (0.000104) z = -2.19	-0.000318*** (0.000105) z = -3.04
Observations	28,975	32,081
Pseudo R-sq	0.1043	0.0960

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 5: Logit model with 'Mixed' fares, estimated using different datasets

Dataset Dependent variable	2011, CAA Hub choice	2011, PaxIS Heathrow Hub choice
Value of airtime	-0.0112*** (0.00215) z = -5.19	-0.0113*** (0.00221) z = -5.11
Empirical frequency	-0.0660*** (0.00335) z = -19.66	-0.0614*** (0.00314) z = -19.52
'Mixed' fare	-0.000644*** (0.000102) z = -6.28	-0.000779*** (0.000105) z = -7.40
Observations	40,627	44,716
Pseudo R-sq	0.2126	0.2047

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Should a fare variable be included?

We have tested the model specification with three different fare variables – 'Synthetic', 'Observed' and 'Mixed' fare data.

Intuitively, we would expect fares to be an important variable in determining passenger choice. From our analysis, we find that the 'Synthetic' data behaves badly in the model – in most cases it enters the equation with the incorrect sign (i.e. positive), or that it is insignificant, as shown in Table 3. In contrast, 'Observed' and 'Mixed' fare variables have significant coefficients and are correctly signed (i.e. negative) in all our models.

In principle this means that either 'Observed' or 'Mixed' datasets could be used. On balance, and notwithstanding the shortcomings of the synthetic fare data referred to above, our preference is to use the mixed fare dataset because the logit model that uses this data provides a better fit (as measured by the "pseudo

R-squared" statistic); and the dataset should be richer by virtue of how it includes fare observations that were not chosen. However, we note that the choice of dataset is a matter of judgment and should also be informed by other factors such as an assessment of the underlying robustness of the data.

Which dataset should be used?

The choice of dataset should be informed by the answers to two questions:

i. does it matter whether we use 2008 or 2011 data? We therefore carried out a 'Seemingly unrelated estimation' (Suest)[9] test in order to test the null hypothesis that the coefficients estimated with 2008 data are similar to the coefficients estimated with 2011 data; and

ii. should CAA or PaxIS data be used? We therefore performed another Suest test to test whether the coefficients estimated with CAA and PaxIS data are similar.

Both tests were conducted on the models without fares.

Table 6 below shows the summary test statistics for the first of our Suest tests.

Table 6: Suest test to compare models without fares, using 2008 and 2011 data

Dataset	Null hypothesis	Chi-sq statistic	Critical Chi-sq at 95%
CAA data	$\beta_{2008} = \beta_{2011}$	28.53	5.991
PaxIS Heathrow data		23.87	

As both Chi-sq test statistics are greater than the critical value at the 95% confidence level, we can reject the null hypothesis that the coefficients estimated with 2008 data are equal to the coefficients estimated with 2011 data.

To answer the second question, we performed a similar Suest test, but the null hypothesis in this case is that the coefficients estimated with CAA data are equal to the coefficients estimated with PaxIS data.

Table 7 below shows the results for the second of our Suest test, where the null hypothesis is instead that the coefficients estimated with CAA and PaxIS data are the same, tested on both 2008 and 2011 data.

Table 7: Suest test for estimating models without fares, with CAA and PaxIS Heathrow data

Dataset	Null hypothesis	Chi-sq statistic	Critical Chi-sq at 95%
2008 data	$\beta_{CAA} = \beta_{PaxIS}$	38.36	5.991
2011 data		80.83	

Again, as both Chi-sq test statistics are greater than the critical value of 95%, we can reject the null hypothesis that the coefficients estimated with CAA data are equal to the coefficients estimated with PaxIS Heathrow data.

Based on these results, we concluded that:

- the relationships between route attributes and their impact on route choice have changed over this time period. In the absence of any clear understanding of why these changes have occurred and what

[9] We have used a specific form of the Wald test called 'Seemingly unrelated estimation' test (Suest test). This takes into account the fact that the factors may not be independent. For further details see pp 264 of "Stata User's Guide release 13".

changes might take place over the timeframe being considered by the Airports Commission, our inclination would be to base the model on the most contemporaneous data and analysis that is available (i.e. the 2011 data); and

- the coefficients estimated using the two datasets are statistically different to each other which suggests that it may be important to make a choice of one data source over the other. However, we have no *a priori* or statistical reason to suggest that one dataset should perform better than the other. Therefore, we recommend that the AC due diligence the two datasets, taking into account factors such as the integrity of the underlying data, to inform a final assessment of its preferred data source.