22nd ATRS World Conference | Seoul 2018

AIRLINES' ENTRY DECISIONS UNDER SCHEDULE OPERATIONAL CONSTRAINTS

Sebastian Birolini[a], Mattia Cattaneo[a], Paolo Malighetti[a], Chiara Morlotti[a,b], Renato Redondi[a]

a. University of Bergamo – Department of Management, Information and Production Engineering
b. University of Pavia – Department of Economics and Management

Air Transport Research Society

INTRODUCTION

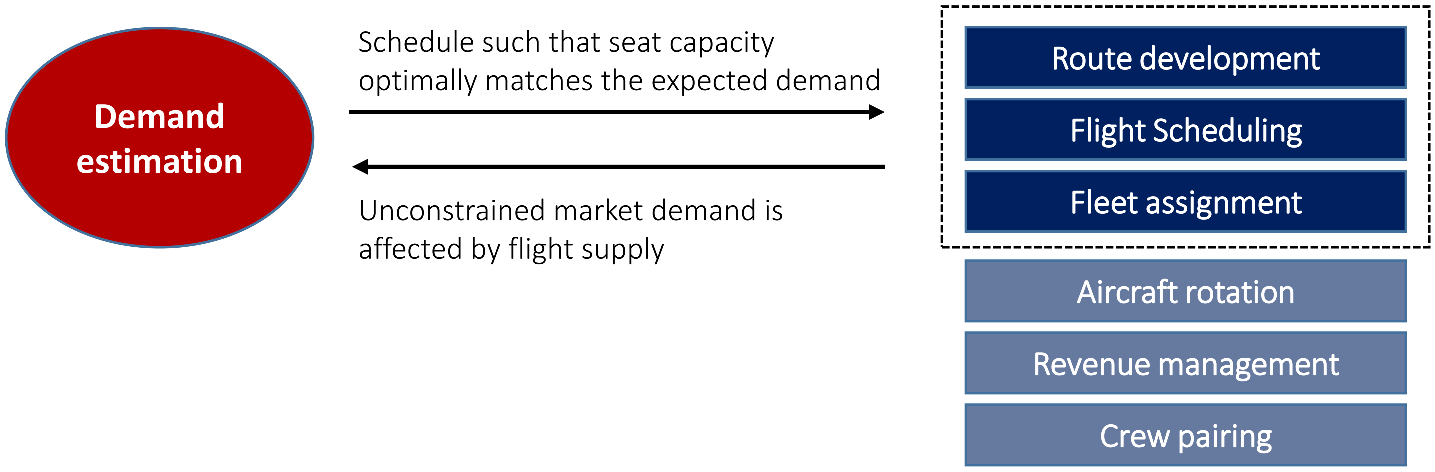
OBJECTIVE

The purpose of the papar is to develop an integrated fleet assignment and schedule planning model which directly accounts for air trip generation over city pairs as a tool to assist airlines' entry and/or re-fleeting decisions.

RATIONALE FOR THE STUDY

> the **interdependency between demand and supply** is a crucial element of airline schedule planning which is not fully taken into account by sequential airline planning schemes

> Schedule evaluation models used for demand estimation provided by private company (e.g. Sabre)



LITERATURE

- Scholars have been developing integrated models that simultaneously optimize the process of selecting flight legs to include in the schedule and assigning aircraft types to these legs.
- Common approaches involve the estimation of **unconstrained demand** in each market and differ in the estimation of passenger flow redistribution across available alternatives:
  - Models based on a **Spill and recapture** process (*Lohatepanont & Barnhart, 2004; Pita et al., 2012; Pita et al., 2014; Sherali et al., 2010, 2013*)
  - Models based on a **Discrete choice modelling** process (*Dong et al., 2016*)

- We contribute to the current literature in a twofold way:
  - **Demand estimation:** by explicitly assessing the impact of itinerary's connectivity and service level on additional demand generation.
  - **Airline planning**: we propose a new methodology to tackle the demand/supply interaction within schedule optimization models leveraging on the estimated impact of newly operated connections.

METODOLOGY

Demand estimation

Development of a **gravity model** (Grosche et al. (2007), Shen (2004), Hwang and Shiao (2011), Srinidhi (2009)) to explore the determinants of air travel demand over city-pairs

Modal split

Definition of a modal split rule to allocate passenger flows over itineraries.

Flight scheduling and fleet assignment

Development of a **Mixed Integer Linear Problem (MILP)** for the integrated flight scheduling and fleet assignment model, where potential demand for each market is estimated simultaneously within the model

DEMAND ESTIMATION: MODEL FORMULATION

- In order to estimate air passenger volume between *city-pairs*, we use the following **gravity model** formulation:

$$D_m = \frac{Pop_o Pop_d}{e^{I_m}} \quad (1)$$

- Which can be linearized through logarithms and estimated by OLS regression:

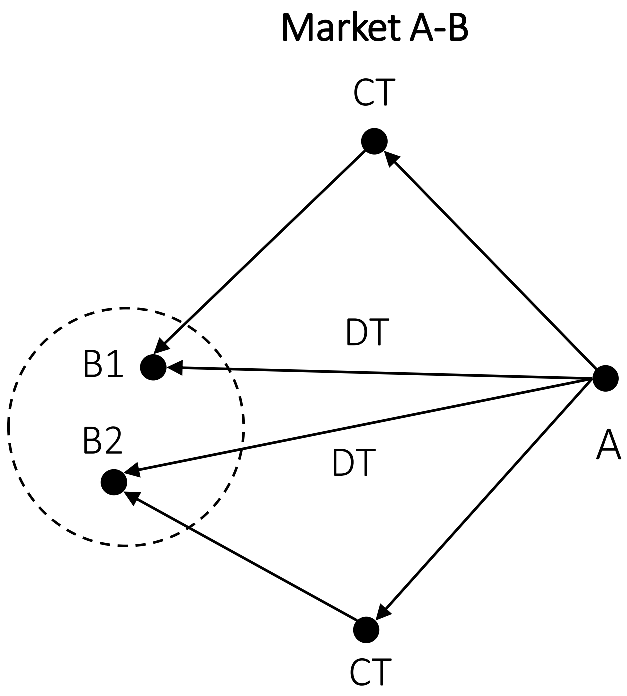
$$ln(D_m) = \beta_{const} + \beta_{Pop_o} ln(Pop_o) + \beta_{Pop_d} ln(Pop_d) + I_m + \varepsilon_m, \quad (2)$$

  - Pop_o: Population at origin
  - Pop_d: Population at destination
  - I_m: Impedance measure (I_m) as function of **distance** and **frequency** (Doganis, 2004) broken down by itinerary type *c*:

$$I_m = \beta_{dist} dist + \sum_{c \in itinerary\_type} \beta_c n_c \quad (3)$$

DEMAND ESTIMATION: CLUSTER ANALYSIS

- To capture the difference contribution to demand generation from itineraries with different characteristics, we perform a clustering analysis by **K-means** algorithm on two features:
  - the **routing factor (RF):** Total Flight Time (TFT) / Direct Flight Time (DT)
  - the **connectivity index (CI)**: Connecting Time (CT) / total travel time (TT)



Itinerary type		Routing factor	Connectivity index
Cluster 0	Nonstop flights	1	0
Cluster 1	Low RF – Low CI	1.50 (*0.22*)	0.33 (*0.07*)
Cluster 2	High RF – Low CI	2.38 (*0.30*)	0.32 (*0.08*)
Cluster 3	Low RF – High CI	1.50 (*0.58*)	0.58 (*0.07*)
Cluster 4	Low RF – High CI	2.28 (*0.30*)	0.57 (*0.07*)

DEMAND ESTIMATION: DESCRIPTIVE STATISTICS

- Data source: OAG, *SEDAC, Columbia University*
- EU top 100 airports by passenger traffic: 290,780 direct flights, 3,326,316 indirect flights
- Indirect itineraries: 1 hour ≤ CT ≤ 6 hours

Variables at market (city-pair) level		**Descriptive Statistics**			
		Mean	*Std. Dev.*	*Min*	*Max*
D	Average passenger flow (daily)	256.78	431.76	0.03	4,275
Pop(o)	Population at origin (within 20 km)	125,880	156,738	578	748,708
Pop(d)	Population at destination (within 20 km)	135,129	164,257	731	748,708
n_itineraries (0)	Average number of direct itineraries (daily)	3.06	4.68	0.03	52.57
n_itineraries (1)	Average number of indirect itineraries (daily) belonging to different clusters	10.72	13.01	0.00	158.00
n_itineraries (2)		5.01	7.98	0.00	76.53
n_itineraries (3)		12.73	17.40	0.00	224.60
n_itineraries(4)		6.92	14.53	0.00	254.63
Distance	Average distance	1,698	633	156	5,230

DEMAND ESTIMATION: MODEL RESULTS

- Coefficients have the expected sign and are statistically significant, except for type-4 itineraries
- Consistently, itineraries with lower routing factor and connectivity index have a greater impact on demand stimulation
- The non-significance of type 4 itineraries' coefficient denotes how low attractive connections (high RF – high CI) do not influence demand
- Distance has a negative impact on passenger flows, although its coefficient is fairly low (intra-EU flights)

Parameters	*Estimate*	*Std.dev*	
ln(Pop_o)	0.0896	(0.0142)	***
ln(Pop_d)	0.0834	(0.0152)	***
n_itineraries (0)	0.1450	(0.0039)	***
n_itineraries (1)	0.0150	(0.0022)	***
n_itineraries (2)	0.0067	(0.0033)	**
n_itineraries (3)	0.0044	(0.0016)	***
n_itineraries (4)	7.54e-04	(0.0018)	
distance	-5.03e-05	(2.70e-05)	*
constant	2.370	(0.2580)	***

Observations	2,962
R-squared	0.563

MODAL SPLIT: ESTIMATION OF ITINERARY DEMAND (1/2)

> Let α be the weighted current provision of air transport services in market m

$$\alpha = \sum_{c \in itinerary\_type} \hat{\beta}_c \bar{n}_c \tag{4}$$

> Let γ represent the following constant for market m:

$$\gamma = e^{\hat{\beta}_{const}} \overline{Pop_d}^{\hat{\beta}_{Pop_d}} \overline{Pop_d}^{\hat{\beta}_{Pop_d}} e^{\hat{\beta}_{dist} dist} \tag{5}$$

> The expected market demand following the introduction of $n_c, c \in itinerary\_types$ new itineraries in market m can be written as follows:

$$\hat{D}' = \gamma e^{\alpha + \sum_{c \in itinerary\_type} \hat{\beta}_c n_c} \tag{6}$$

ASS1: Itineraries with same characteristics get the same market share

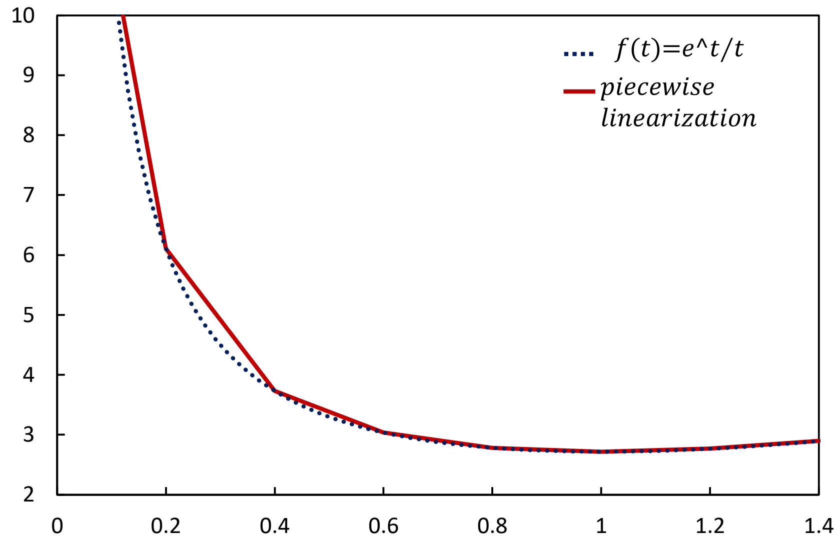
ASS2: Potential demand on itinerary-type c is proportional to estimated beta coefficients from previous regression

MODAL SPLIT: ESTIMATION OF ITINERARY DEMAND (2/2)

- Holding the assumptions true and setting $t = \alpha + \sum_{c \in itinerary\_type} \hat{\beta}_c n_c$ the demand for each itinerary type c in market can be estimated as follows:

$$d'_c = \gamma \frac{e^t}{t} \beta_c \qquad (7)$$

- Which is a univariate function that can be directly entered into our MILP by means of **piecewise linearization**.



OPTIMIZATION MODEL: NOTATION

SETS

$N = set\ of\ activity\ nodes, indexed\ by\ n$

$M = set\ of\ markets, indexed\ by\ m$

$I = set\ of\ itineraries, indexed\ by\ i$

$F = set\ of\ flight\ arcs, indexed\ by\ f$

$F^m \subset F = set\ of\ mandatory\ flight\ arcs$

$F^o \subset F = set\ of\ optional\ flight\ arcs$

$G = set\ of\ ground\ arcs, indexed\ by\ g$

$CL_f = set\ of\ flight\ arcs\ crossing\ the\ count\ line$

$CL_g = set\ of\ ground\ arcs\ crossing\ the\ count\ line$

$AT = set\ of\ fleet\ types$

$F_n^{in} = set\ of\ flight\ arcs\ arriving\ at\ activity\ node\ n$

$F_n^{out} = set\ of\ flight\ arcs\ departing\ from\ activity\ node\ n$

$G_n^{in} = set\ of\ ground\ arcs\ arriving\ at\ activity\ node\ n$

$G_n^{out} = set\ of\ ground\ arcs\ departing\ from\ activity\ node\ n$

$I_m = set\ of\ itineraries\ in\ market\ m$

$F_i = set\ of\ flight\ arcs\ in\ itinerary\ i$

$I_f = set\ of\ itineraries\ featuring\ flight\ arc\ f$

PARAMETERS

$fleet_a = fleet\ size\ by\ aircraft\ type$

$cap_a = passenger\ capacity\ by\ aircraft\ type$

$N_i = number\ of\ flight\ legs\ in\ itinerary\ i$

$\beta_i = beta\ coefficient\ for\ itinerary\ i$

$\alpha_m = current\ weighted\ air\ service\ provision\ in\ market\ m$

$\gamma_m = market\ m's\ constant$

$c_f^{pax} = variable\ cost\ per\ pax\ (landing\ fees, etc)$

$c_f^{fix} = fixed\ cost\ (airport\ fees, etc)$

$c_a^{op} = flight\ costs\ per\ hour\ of\ flight\ for\ an\ aircraft\ of\ type\ a$

$ftime_f = flight\ time\ for\ flight\ arc\ f$

$fare_i = price\ of\ itinerary\ i$

VARIABLES

$x_{a,f} \in 0,1$	$= 1\ aircraft\ type\ a\ is\ assigned\ to\ flight\ arc\ f$ $= 0\ aircraft\ type\ a\ is\ not\ assigned\ to\ flight\ arc\ f$
$y_{a,f} \in N^+$	$the\ flow\ value\ of\ aircraft\ type\ a\ though\ ground\ arc\ g$
$k_i \in 0,1$	$= 1\ itinerary\ i\ is\ operated$ $= 0\ itinerary\ i\ is\ not\ operated$
$t_m \in R^+$	$the\ weighted\ air\ service\ provision\ in\ market\ m$
$d_m \in R^+$	$f(t_m) for\ market\ m\ such\ that\ d_m \gamma_m \beta_i\ gives\ expected\ demand\ for\ itinerary\ i\ in\ market\ m$
$q_i \in N^+$	$the\ flow\ of\ passengers\ accommodated\ on\ itinerary\ i$

OPTIMIZATION MODEL: MODEL FORMULATION

Max

$$\sum_{i\in I} q_i\, fare_i - \sum_{a\in AT}\sum_{f\in F} x_{a,f}\, c_f^{pax} - \sum_{f\in F} c_f^{pax} \sum_{i\in I_f} q_i - \sum_{a\in AT} c_a^{op} \sum_{f\in F} x_{a,f}\, ftime_f$$

Subject to:

(1) Flow balance

$$\sum_{f\in F_n^{in}} x_{a,f} + \sum_{g\in G_n^{in}} y_{a,g} = \sum_{f\in F_n^{out}} x_{a,f} + \sum_{G\in G_n^{out}} y_{a,g}\,, \qquad \forall\, n\in N, \forall\, a \in AT$$

(2) Schedule repeatability

$$\sum_{n\in SL}\sum_{f\in F_n^{out}} x_{a,f} + \sum_{n\in SL}\sum_{g\in G_n^{out}} y_{a,g} = fleet_a, \qquad \forall\, a\in AT$$

(3) Fleet availability

$$\sum_{f\in CL_f} x_{a,f} + \sum_{g\in CL_g} y_{a,g} \le fleet_a, \qquad \forall\, a\in AT$$

(4)-(5) Mandatory/ optional flight legs

$$\sum_{a\in AT} x_{a,f} = 1, \qquad \forall\, f\in F^o$$

$$\sum_{a\in AT} x_{a,f} \le 1, \qquad \forall\, f\in F^m$$

6)-(7) Itinerary status constraints

$$k_i - \sum_{a\in AT} x_{a,f} \le 0, \qquad \forall\, f\in F_i,\ \forall\, i\in I$$

$$k_i - \sum_{a\in AT}\sum_{f\in FA_i} x_{a,f} \ge 1 - N_i, \quad \forall\, i\in I$$

(8)-(9) market demand

$$t_m = \alpha_m + \sum_{i\in I_m} \beta_i k_i, \quad \forall\, m\in M$$

$$d_m = \frac{e^{t_m}}{t_m}, \quad \forall\, m\in M$$

(10)-(11) passengers per itinerary

$$q_i \le \gamma_m d_m \beta_i, \quad \forall\, i\in I$$

$$q_i \le M_{big} k_i, \quad \forall\, i\in I$$

(12) aircraft capacity

$$\sum_{i\in I_f} q_i \le \sum_{a\in AT} x_{a,f}\, cap_a\,, \quad \forall\, f\in F$$

CONCLUSIONS & FUTURE DEVELOPMENTS

- Computational complexity for large-scale problem due to the high number of binary variables
  - Develop *specialized algorithms* to solve larger instances or add additional elements to the model formulation
  - Adopt an *iterative process*
- Consider additional features to be included in the clustering process to better distinguish itinerary types (prices, timetable differentiation)
- Improve the gravity model formulation:
  - Include variables describing the general economic activity and geographical characteristics of city-pairs to better deal with city-pairs where poorly air service is currently established
  - Distinguish among Intercontinental flights and LCCs vs FSCs
  - Test for other functional forms and tackle causality issues

22nd ATRS World Conference | Seoul 2018

AIRLINES' ENTRY DECISIONS UNDER SCHEDULE OPERATIONAL CONSTRAINTS

Sebastian Birolini
sebastian.birolini@unibg.it