AIRLINES’ ENTRY DECISIONS UNDER SCHEDULE OPERATIONAL CONSTRAINTS

Sebastian Birolini\textsuperscript{a}, Mattia Cattaneo\textsuperscript{a}, Paolo Malighetti\textsuperscript{a}, Chiara Morlotti\textsuperscript{a,b}, Renato Redondi\textsuperscript{a}

\textsuperscript{a}. University of Bergamo – Department of Management, Information and Production Engineering
\textsuperscript{b}. University of Pavia – Department of Economics and Management
INTRODUCTION

OBJECTIVE
The purpose of the paper 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 \textit{unconstrained demand} in each market and differ in the estimation of passenger flow redistribution across available alternatives:

> 1. Models based on a \textit{Spill and recapture} process (Lohatepanont & Barnhart, 2004; Pita et al., 2012; Pita et al., 2014; Sherali et al., 2010, 2013)

> 2. Models based on a \textit{Discrete choice modelling} process (Dong et al., 2016)

> We contribute to the current literature in a twofold way:

> 1. **Demand estimation**: by explicitly assessing the impact of itinerary’s connectivity and service level on additional demand generation.

> 2. **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:

\[ D_m = \frac{Pop_o Pop_d}{e^{I_m}} \]  

(1)

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

\[ \ln(D_m) = \beta_{\text{const}} + \beta_{\text{Pop}_o} \ln(Pop_o) + \beta_{\text{Pop}_d} \ln(Pop_d) + I_m + \varepsilon_m \]  

(2)

> \(\text{Pop}_o\): Population at origin
> \(\text{Pop}_d\): Population at destination
> \(I_m\): Impedance measure \(I_m\) as function of distance and frequency (Doganis, 2004) broken down by itinerary type \(c\):

\[ I_m = \beta_{\text{dist dist}} + \sum_{c \in \text{itinerary type}} \beta_c n_c \]  

(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)

<table>
<thead>
<tr>
<th>Itinerary type</th>
<th>Routing factor</th>
<th>Connectivity index</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cluster 0</strong></td>
<td>Nonstop flights</td>
<td>1</td>
</tr>
<tr>
<td><strong>Cluster 1</strong></td>
<td>Low RF – Low CI</td>
<td>1.50 (0.22)</td>
</tr>
<tr>
<td><strong>Cluster 2</strong></td>
<td>High RF – Low CI</td>
<td>2.38 (0.30)</td>
</tr>
<tr>
<td><strong>Cluster 3</strong></td>
<td>Low RF – High CI</td>
<td>1.50 (0.58)</td>
</tr>
<tr>
<td><strong>Cluster 4</strong></td>
<td>Low RF – High CI</td>
<td>2.28 (0.30)</td>
</tr>
</tbody>
</table>
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 \(\leq CT \leq 6\) hours

---

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

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

Observations: 2,962
R-squared: 0.563
MODAL SPLIT: ESTIMATION OF ITINERARY DEMAND (1/2)

> Let $\alpha$ be the weighted current provision of air transport services in market $m$

$$\alpha = \sum_{c \in \text{itinerary\_type}} \hat{\beta}_c \tilde{n}_c \quad (4)$$

> Let $\gamma$ represent the following constant for market $m$:

$$\gamma = e^{\beta_{\text{const}} \text{Pop}_d \hat{\beta}_{\text{Pop}_d} \text{Pop}_d \hat{\beta}_{\text{Pop}_d} e^{\hat{\beta}_{\text{dist}} \text{dist}}} \quad (5)$$

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

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

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

\[
d'_c = \gamma \frac{e^t}{t} \beta_c
\]

(7)

Which is a univariate function that can be directly entered into our MILP by means of piecewise linearization.
## OPTIMIZATION MODEL: NOTATION

### SETS

- $N = \text{set of activity nodes, indexed by } n$
- $M = \text{set of markets, indexed by } m$
- $I = \text{set of itineraries, indexed by } i$
- $F = \text{set of flight arcs, indexed by } f$
- $F_m \subset F = \text{set of mandatory flight arcs}$
- $F^o \subset F = \text{set of optional flight arcs}$
- $G = \text{set of ground arcs, indexed by } g$
- $\text{CL}_f = \text{set of flight arcs crossing the count line}$
- $\text{CL}_g = \text{set of ground arcs crossing the count line}$
- $\text{AT} = \text{set of fleet types}$
- $F^\text{in}_n = \text{set of flight arcs arriving at activity node } n$
- $F^\text{out}_n = \text{set of flight arcs departing from activity node } n$
- $G^\text{in}_n = \text{set of ground arcs arriving at activity node } n$
- $G^\text{out}_n = \text{set of ground arcs departing from activity node } n$
- $I_m = \text{set of itineraries in market } m$
- $F_i = \text{set of flight arcs in itinerary } i$
- $I_f = \text{set of itineraries featuring flight arc } f$

### PARAMETERS

- $\text{fleet}_a = \text{fleet size by aircraft type}$
- $\text{cap}_a = \text{passenger capacity by aircraft type}$
- $N_i = \text{number of flight legs in itinerary } i$
- $\beta_i = \text{beta coefficient for itinerary } i$
- $\alpha_m = \text{current weighted air service provision in market } m$
- $\gamma_m = \text{market } m\text{'s constant}$
- $c^\text{pax}_f = \text{variable cost per pax (landing fees, etc)}$
- $c^\text{fix}_f = \text{fixed cost (airport fees, etc)}$
- $c^\text{op}_a = \text{flight costs per hour of flight for an aircraft of type } a$
- $\text{ftime}_f = \text{flight time for flight arc } f$
- $\text{fare}_i = \text{price of itinerary } i$

### VARIABLES

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

Max

\[
\sum_{i \in I} q_i \text{fare}_i - \sum_{a \in AT} \sum_{f \in F} x_{a,f} pax_{f} - \sum_{f \in F} c_{f}^{\text{pax}} \sum_{i \in I_f} q_i - \sum_{a \in AT} c_{a}^{\text{op}} \sum_{f \in F} x_{a,f} \text{ftime}_f
\]

Subject to:

(1) Flow balance

\[
\sum_{f \in F_{n}^\text{in}} x_{a,f} + \sum_{g \in G_{n}^\text{in}} y_{a,g} = \sum_{f \in F_{n}^\text{out}} x_{a,f} + \sum_{g \in G_{n}^\text{out}} y_{a,g}, \quad \forall n \in N, \forall a \in AT
\]

(2) Schedule repeatability

\[
\sum_{n \in SL} \sum_{f \in F_{n}^\text{out}} x_{a,f} + \sum_{n \in SL} \sum_{g \in G_{n}^\text{out}} y_{a,g} = \text{fleet}_a, \quad \forall a \in AT
\]

(3) Fleet availability

\[
\sum_{f \in CL_f} x_{a,f} + \sum_{g \in CL_g} y_{a,g} \leq \text{fleet}_a, \quad \forall a \in AT
\]

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

\[
\sum_{a \in AT} x_{a,f} = 1, \quad \forall f \in F^o
\]

\[
\sum_{a \in AT} x_{a,f} \leq 1, \quad \forall f \in F^m
\]

6)-(7) Itinerary status constraints

\[
k_i - \sum_{a \in AT} x_{a,f} \leq 0, \quad \forall f \in F_i, \forall i \in I
\]

\[
k_i - \sum_{a \in AT} \sum_{f \in FA_i} x_{a,f} \geq 1-N_i, \quad \forall i \in I
\]

(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

\[
q_i \leq \gamma_m d_m \beta_i, \quad \forall i \in I
\]

\[
q_i \leq M_{big} k_i, \quad \forall i \in I
\]

(12) Aircraft capacity

\[
\sum_{i \in I_f} q_i \leq \sum_{a \in AT} x_{a,f} \text{cap}_a, \quad \forall f \in F
\]
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
AIRLINES’ ENTRY DECISIONS UNDER SCHEDULE OPERATIONAL CONSTRAINTS

Sebastian Birolini
sebastian.birolini@unibg.it