A bilevel modelling approach to pricing and fare optimisation in the airline industry

Jean-Philippe Côté, Patrice Marcotte* and Gilles Savard
Received (in revised form): 21st March, 2003

*Département d'informatique et de recherche opérationnelle and Centre de recherches sur les transports, Université de Montréal, CP 6128, succursale Centre-Ville, Montréal, Québec, Canada H3C 3J7
Tel: +1 514 343 5941; Fax: +1 514 343 7121; E-mail: marcotte@iro.umontreal.ca

Jean-Philippe Côté *holds a BSc in mathematics–computer science from McGill University and a Masters in Operations Research from the University of Montreal. He acquired practical experience at major software corporations and worked as a research analyst. He is currently engaged in a PhD programme where his work involves the development of efficient algorithms for solving large-scale bilevel pricing problems.*

Patrice Marcotte *has been Professor of Operations Research at the Computer Science and Operations Research Department of the University of Montreal for the last nine years. As an expert in transportation modelling and mathematical programming, he has published more than sixty papers in international journals. He is currently associate editor of the journals* Operations Research, Transportation Science, Operations Research Letters, Networks and Spatial Economics *and* Advances in Nonlinear Variational Inequalities. *He stands on the board of the Center for Research on Transportation of the University of Montreal.*

Gilles Savard *is Head of the Department of Mathematics and Industrial Engineering at École Polytechnique de Montréal and a member of the board of direction of GERAD (Groupe d'études et de recherche en analyse des décisions). He has published more than forty papers in major refereed scientific journals. He is a leading expert in both the theory and practice of bilevel programming and in the development of global algorithms for nonconvex optimisation. He spent one year as project director in a firm specialising in software for the airline industry.*

Abstract

KEYWORDS: airline industry, revenue management, yield management, pricing decision support system, bilevel programming, integer programming

The airline revenue management problem can be decomposed into four distinct but related sub-problems that are usually treated separately: demand forecasting, overbooking, capacity allocation and pricing. In recent decades, much interest has been devoted to overbooking and capacity allocation issues and, today, most major airlines rely on computerised tools to deal with these two sub-problems. Pricing, however, has received less attention, which can be explained by the technical and theoretical difficulties inherent in the implementation of a practical pricing decision support system. This paper presents a new modelling approach which

Journal of Revenue and Pricing Management, Vol. 2, No. 1, 2003, pp. 23–36

allows for the joint solution of the capacity allocation and pricing sub-problems faced by a major North American airline. Using pre-defined booking limits, the resulting model can also be applied in a 'pure' pricing context. This approach is based on the bilevel programming paradigm, a special case of hierarchical mathematical optimisation. This modelling technique makes it possible to take into account matters such as customer segmentation, behaviour with regard to fares and other product attributes, and the interactions induced by overlapping routes, which are typical of modern air transportation networks.

INTRODUCTION

It is generally agreed (see for instance Kimes, 1990) that airline revenue management can be decomposed into four sub-problems: demand forecasting, overbooking policy determination, capacity allocation (sometimes called seat inventory control) and pricing. These four interdependent issues are the links which form the revenue optimisation chain of an airline. While demand forecasting remains a fundamentally statistical task, the determination of overbooking and capacity allocation policies is usually addressed using optimisation techniques. Indeed, these issues have been the subject of several studies in the operations research or management science literature, over the last thirty years. As a result, major airlines nowadays rely on computerised tools for addressing overbooking and seat allocation problems. Yet while experts agree that pricing/fare optimisation is a vital ingredient of the revenue management process (see Kretsch, 1995; Garvett and Michaels, 1998), few papers deal with this topic. The reasons for this situation are twofold. First, while a comprehensive pricing model must involve stochastic, dynamic and game theoretic elements, there is no agreed modelling approach within the revenue management community. Next, the gathering of data required in a real-life application is a daunting task. One must, for example, track down information about competitors' fares, demand forecast, historical sales patterns, etc. Airlines generally rely on computerised database systems to store such information. As these systems seldom use compatible formats, one must extract, interpret and reformat the required data in order to establish links between the various sources of information and develop a coherent database. Refer to Kretsch (1995) for an overview of the difficulties surrounding the retrieval and treatment of fare data.

This paper introduces a novel methodology for solving the pricing/fare optimisation problem. The modelling approach is based on the bilevel mathematical programming paradigm, a special case of hierarchical mathematical optimisation. The resulting mathematical model, which is both static and deterministic, distinguishes itself from alternative approaches in important ways:

- disaggregated and O&D-based representation of demand, which allows for accurate customer segmentation;
- modelling of customer reaction to fares and flight attributes (duration, quality of service, fidelity programmes, weekend restrictions, etc.);
- prices and capacity allocations set at fare basis code[1] level;
- full accounting of the schedules and networks of the client airline and its competitors.

The decision variables of the mathematical model are divided into two subsets: product fares and number of passengers buying a given product. Conceptually, the model does not require any additional information and will generate not only optimal fare schedules, but also the corresponding optimal seat allocation and booking class limits. If, however, booking limits on some or all fare classes are imposed, the model can be operated in 'pure' pricing mode, complying with the booking class constraints. It must be realised that these constraints greatly restrict the model's flexibility, and might lead to inferior revenues. Whenever this is feasible, one should let the model decide by itself the optimal number of seats to allow in each booking class. For a given booking class, this number is simply the sum of the number of passengers buying fare products belonging to this class. As to the fare variables, they may be set free, ie subject to no restrictions whatsoever.

The rest of this paper is organised as follows. The second section gives a short overview of the scientific literature on revenue management and pricing in the airline industry. The third section introduces the bilevel programming paradigm, describes the model and provides an illustrative example. Practical issues related to the use of this approach in a real-life context are revealed in the fourth section, while the final section proposes extensions to the current model.

LITERATURE REVIEW

A recent survey of revenue management can be found in the work of McGill and Van Ryzin (1995). In the airline industry, it can be characterised in several ways. Weatherford and Bodily (1992) adopt the mathematical programming point of view and classify the various objectives to be optimised (revenue, capacity utilisation, customer utility), as well as the operational, financial and marketing constraints. They propose a taxonomy of revenue management models based on the objectives that are explicitly addressed and the constraints that must be satisfied. A unified approach seeking to deal simultaneously with all aspects of the problem has, however, been judged intractable in practice by Smith *et al.* (1992). Their paper, based on the experience acquired at a major American carrier, suggests that a decomposition of the general problem into structurally simpler sub-problems is unavoidable, given today's theoretical and technological capabilities. Kimes (1990) suggests the decomposition introduced earlier, ie demand forecasting, overbooking policy determination, seat allocation and pricing.

This work is primarily concerned with the *joint* seat allocation and pricing processes. Seat inventory control can be traced back to the work of Littlewood (1972). One can distinguish between so-called 'static' models, where bookings are taken sequentially, starting with the lowest class and proceeding to the next class only when the current class is filled, and 'dynamic' methods, which make no assumption on the order of arrival of booking requests. There is a further distinction between leg-based methods, which ignore the interactions between flight legs, and broader, network-based methods. For articles dealing with static, leg-based approaches, see Brumelle *et al.* (1990) and Belobaba (1987a,b, 1998). This last author has proposed heuristics usually referred to as $EMSR_a$ and $EMSR_b$, which are often used in practice. Optimality conditions for such static methods have been obtained by Curry (1990), Brumelle and McGill (1993), Wollmer (1992) and Robinson (1993). Dynamic versions of the leg-based methods have been proposed, most of them relying on dynamic programming for their solution. Examples of such approaches can be found in Lee and Hersh (1993), Subramanian *et al.* (1999), Lautenbacher and Shaler (1999) and Feng and Xiao (2001). They are differentiated, in particular, by the consideration (or not) of no-shows and cancellations.

Network-based seat inventory control methods can be divided into two broad categories. Improved versions of leg-based methods that seek to exploit some of the network structure through modelling devices such as virtual booking classes fall into the first category. The main advantage of these approaches is that they exploit the network structure and, since they rely on existing frameworks, require small investment. See Belobaba (1998) and Smith *et al.* (1992) for further details. The second category, which regroups methods that are explicitly based on linear programming network formulations, includes the works of Glover *et al.* (1992), Curry (1990) and Wong *et al.* (1993). In De Boer *et al.* (1999), deterministic and stochastic network formulations are compared, and numerical results are presented. Surprisingly, these authors have found that simple deterministic models based on expected demand generate better revenue than more sophisticated probabilistic models. According to Belobaba (1998), these methods can take into account the combinatorial structure inherent in any real-life transportation system. They are not without drawbacks, however. First, the integration of dynamic and stochastic aspects within a network formulation can only be achieved at the expense of a significant increase in complexity. Secondly, the seat allocation schemes they produce are non-nested and highly 'partitioned', ie the 'optimal' number of seats allocated to a given booking class on a given itinerary may be too small to be of any practical use. Nevertheless, linear programming network formulations are sometimes used as an auxiliary procedure in 'bid price' methods. Talluri and van Ryzin (1998) describe bid prices as threshold values set for each leg of a network, such that a product (which can be seen as an itinerary in the network) is sold only if its fare exceeds the sum of the bid prices of the legs along its path. In other words, the bid price of a leg represents the marginal contribution of this leg to the global network revenue. Those familiar with linear programming will recognise that this is the classical interpretation of dual multipliers of capacity constraints in a capacitated maximum-flow network formulation, hence the use of such models in bid price methods. More details can be found in Belobaba (1998), Pak and Piersma (2002), Talluri (2001), Talluri and van Ryzin (1998) and Vinod (1995).

The literature on airline pricing and fare optimisation is much more limited. Gallego and Van Ryzin (1994, 1997) and You (1999) model the pricing of multi-leg flights as a stochastic decision process, and solve the resulting recurrence equations by dynamic programming techniques. They underline the strong duality relationship between pricing and seat allocation and argue that the latter can be derived from the former by allowing prices to become infinite on a given booking class, which is tantamount to closing it. Recently, Kuyumcu and Garcia-Diaz (2000) claimed that their model jointly solved the pricing and seat allocation problems. This appears to be an overstatement. The authors actually formulate and solve (exactly or heuristically) an integer program model which, given a restricted number of fare structures, determines the one that should be implemented, together with the number of seats to be allocated to each class. Fares are therefore exogenous data, and not determined by the optimisation process.

A BILEVEL MODEL OF FARE OPTIMISATION

Bilevel programming

Bilevel programming is a branch of hierarchical mathematical optimisation in which an agent, referred to as the 'leader', integrates within its optimisation process the reaction of a rational 'follower' to his decisions. Once the leader has set its decision variables, the follower solves an optimisation problem, taking as exogenous the leader's variables. Such a model is relevant whenever the optimiser controls only indirectly the choices of agents affected by its decisions, and is pervasive in a free market.

In this sequential framework, the leader has full knowledge of the follower's optimisation problem. For a given leader vector x, the follower solves an optimisation problem where both its objective function $f(x,y)$ and its feasible set may be dependent on x. The general form of a non-linear bilevel program is

$$\begin{aligned} \max_{x,y} \quad & F(x,y) \\ \text{s.t.} \quad & (x,y) \in X \\ & y \in \arg\min_{y' \in Y(x)} f(x,y'). \end{aligned}$$

With a slight abuse of notation, the model is presented in the so-called vertical format

$$\max_{x,y} F(x,y) \tag{1}$$

$$\text{s.t.}\ (x,y) \in X \tag{2}$$

$$\min_{y} f(x,y) \tag{3}$$

$$\text{s.t.}\ y \in Y(x). \tag{4}$$

If, for given x, the reaction $y(x)$ is unique and available in closed form, then the bilevel program can be rewritten as the single-level program

$$\begin{aligned} \max_{x} \quad & F(x,y(x)) \\ \text{s.t.} \quad & (x,y(x)) \in X. \end{aligned}$$

Such reformulation, however, hides the features of the lower-level problems and is only applicable in problems of little practical interest.

In the present case, the leader is the client airline. Its objective is to maximise the revenue raised on the entire network over a given time horizon. Revenue is obtained by summing the products of fares and number of passengers (or 'flow') buying the corresponding fare product. While fares are under the direct control of the leader, the number of buyers in each fare class (lower level) is dependent on fare structure, seat availability, flight duration, etc. At the lower level, users minimise their own disutility of travel from their origin point to their destination point. Decisions taken by the client airline at the upper level are parameters that are integrated in the objective function and constraints of the customers.

In the model, the perceived travel disutilities (PTDs) are linear combinations of flight attributes, which include fare and flight duration, of course, but also quality of service (cabin), flight restrictions (max-stay, min-stay, Saturday night rule), etc. Distinct passengers may have different valuations of the criteria, ie PTDs are not uniform throughout the population. Actually, customers are partitioned into 'user groups', each group being endowed with the same valuation of the products. For instance, in the simple situation where only two criteria are considered (fare and duration, say), passengers within a group assign the same value to one unit of travel time.

Model background

This section introduces the concepts and assumptions underlying the model. Throughout the paper, the term 'flight' is used to denote the journey completed by a traveller from his origin to his final destination. A flight is composed of one or several 'legs', which are uninterrupted rides of an aircraft that take place between take-off and landing. A direct flight is composed of a single leg, while a connecting flight has two or more legs separated by sojourns at transit airports. The term 'market' refers to a pair of cities (origin and destination) for which a flight is sold.

It is assumed that demand for each user group and each market is available and that the fare structure of all airlines, both the leader and the competition, can be accessed in real time. The conservative assumption that competing flights have unlimited capacity is also made. This hypothesis is common to most pricing models and is generally satisfied for most markets. It may be justified by the high degree of competitiveness in the airline industry and, on the pragmatic side, by the difficulty of obtaining accurate capacity data for *each* booking class; if such data were available, this assumption could be lifted and capacity constraints readily integrated in the model. Based on the above information, the model assigns demand to the products of the leader and those of the competition, for each market and each user group.

The model is driven by parameters that fall into two distinct categories. The first category consists of quantities that describe the purchasing behaviour of each user group. As stated previously, passengers base their travel decisions on the valuation of attributes that characterise the products. The foremost criterion is the nominal value of an airline ticket. Other criteria include flight duration; quality of service (which can be seen as the aggregation of a product's attributes, such as cabin, minimum sojourn time, maximal sojourn time, Saturday sleep-over restrictions, refundable ticket, etc.); customer inertia (fidelity to a given airline resulting from a rewards programme), etc. The behavioural parameters allow the value of these criteria for the users of each group to be quantified in monetary units. They yield group disutilities, which are weighted sums of a product's attributes.

The second category of parameters is associated with commercial targets set by the leader airline, such as revenue or market targets. As an example, the leader, for marketing or visibility reasons, may wish to increase its passenger share from 25 to 40 per cent on some strategic market. The model, through its bilevel structure, will then adjust its fare structure (most likely downwards, but not necessarily) so as to induce a traffic increase on that market. It will then find the best solution that meets the market share targets.

Furthermore, the model uses exogenous quantities related to physical constraints (aircraft capacities) or rules set by regulatory agencies or by governments. For instance, lower bounds on fares could be imposed to prevent predatory behaviour from a dominant airline. Symmetrically, fares on small monopolistic regional markets may be subject to ceilings.

As mentioned earlier, the model has the capability of jointly solving the pricing and seat allocation problems. Fares correspond to decision variables and constitute its primary output, while booking limits can be computed by summing up the values of the flow variables over all fare products belonging to a class. However, for technical and managerial reasons, the authors' approach is likely to find its use as a pure pricing tool; this implies that it must be able to deal with exogenous, *a priori* booking limits, through their integration in the model formulation.

Notation and mathematical formulation

In this section, the bilevel formulation of the model is set out, restricted to three flight attributes: fare, duration and quality of service. Throughout, superscript '1' refers to the leader airline, and superscript '2' to the aggregated competition. Sets are represented in capital Latin letters. The sets and corresponding indices are:

k	market index
f	flight index
s	leg index
p	fare product index
b	booking class index
g	user group index
K	set of all markets (origin-destination pairs)
F^1	set of flights offered by the leader

F^2	set of flights offered by the competition
$F^i(k)$	set of flights offered by agent i on market k: $F^i = \cup_k F^i(k)$
S	set of flight legs operated by the leader airline
$S(f) \subset S$	set of flight legs making flight $f \in F^1$
$P(f)$	set of fare products (corresponding to fare basis codes) offered on flight f
$B(f)$	set of booking classes open on flight f
G	set of user groups
$b(p)$	booking class of the product $p \in P(f)$.[2]

Fares and flight attributes are:

$t^1_{p,f}$	fare of the leader airline for product p on flight f (decision variable)
$t^2_{p,f}$	fare of the competition for product p on flight f (exogenous data)
t^i	fare vector
DUR_f	duration of flight f
$QOS_{p,f}$	quality of service associated with product p on flight f.

At the lower level, decision variables are passenger flows:

$v^i_{p,f,g}$	number of passengers of group $g \in G$ purchasing product $p \in P(f)$ on flight $f \in F^i$.

Lower case Greek letters denote the parameters of the model, which, as mentioned earlier, either characterise user behaviour or are related to commercial goals. The behavioural parameters α and β denote, respectively, the valuation of one unit of duration and one unit of quality of service. For instance, if a given flight has a duration of 140 minutes and passengers of a given group have a valuation of time of α = \$2.5/min., then the contribution of the duration criterion to the PTD is 140 min $\times$ \$2.5/min = \$350. The following are set:

$\alpha_{g,k}$	monetary equivalent of one duration unit for passengers of group g on market k
$\beta_{g,k}$	monetary equivalent of one unit of disutility for passengers of group g on market k
$c^i_{p,f,g}(\alpha,\beta)$	perceived travel disutility

$$c^i_{p,f,g}(\alpha,\beta) = t^i_{p,f} + \alpha_{g,k} \times DUR_f + \beta_{g,k} \times QOS_{p,f}$$

$\underline{\sigma}_k$:	lower bound on targeted flow of market k
$\overline{\sigma}_k$:	upper bound on targeted flow of market k
$\underline{\rho}_k$:	lower bound on targeted revenue of market k
$\overline{\rho}_k$:	upper bound on targeted revenue of market k.

Besides the parameters listed above, the model requires the following input:

d_k:	total demand on market k over the planning horizon
$h_{g,k}$:	fraction of demand d_k belonging to group g
$l_{b,f}$:	number of seats available in class b on leader flight f (booking limit)
u_s:	aircraft capacity of flight segment s.

Each element of the mathematical formulation is now discussed. At the upper level of the bilevel program, the leader maximises revenue over the entire network, ie by considering all its flights. This yields the objective

$$\max_{t^1,v^1,v^2} \sum_{f\in F^1} \sum_{p\in P(f)} \sum_{g\in G} t^1_{p,f}\, v^1_{p,f,g}$$

At the lower level, passengers meet their travel demand at least cost, ie they minimise their individual PTD. Since there is no interrelation between passenger flows, this is tantamount to minimising the disutility of the entire population:

$$\min_{v^1, v^2} \sum_{f \in F^1} \sum_{p \in P(f)} \sum_{g \in G} c^1_{p,f,g}(\alpha, \beta)\, v^1_{p,f,g} + \sum_{f \in F^2} \sum_{p \in P(f)} \sum_{g \in G} c^2_{p,f,g}(\alpha, \beta)\, v^2_{p,f,g}$$

Flows produced by the model must respect targets set by the commercial parameters, as well as physical booking limits. This yields the passenger and revenue share constraints

$$\underline{\sigma}_k \leq \sum_{f \in F^1(k)} \sum_{p \in P(f)} \sum_{g \in G} v^1_{p,f,g} \leq \overline{\sigma}_k \quad \forall k \in K$$

$$\underline{\rho}_k \leq \sum_{f \in F^1(k)} \sum_{p \in P(f)} \sum_{g \in G} v^1_{p,f,g}\, t^1_{p,f} \leq \overline{\rho}_k \quad \forall k \in K$$

$$\sum_{p \in P(f) | b(p)=b} \sum_{g \in G} v^1_{p,f,g} \leq l_{b,f} \quad \forall b \in B(f),\ \forall f \in F^1$$

If the model is used to solve the pricing and seat allocation subproblems jointly, ie without taking into account any *a priori* booking limits, the capacity constraint takes the form

$$\sum_{f | s \in S(f)} \sum_{p \in P(f)} \sum_{g \in G} v^1_{p,f,g} \leq u_s \qquad \forall s \in S$$

Note that, in this latter case, booking limits are obtained *a posteriori* through the summation

$$l_{b,f} = \sum_{p \in P(f) | b(p)=b} \sum_{g \in G} v^1_{p,f,g} \quad \forall f \in F^1, \quad \forall b \in B(f)$$

Next, demand must be satisfied for each user group and each market:

$$\sum_{f \in F^1} \sum_{p \in P(f)} v^1_{p,f,g} + \sum_{f \in F^2} \sum_{p \in P(f)} v^2_{p,f,g} = d_k\, h_{g,k} \qquad \forall g \in G,\ \forall k \in K$$

In the model, one decision variable is assigned to each flight/fare product combination. Most fare filing systems, however, allow only for the publication of fare basis codes on a per-market basis, without associating them with a particular flight. As a consequence, an available product, say unrestricted full economy fare, can be purchased on several flights scheduled on a market, no matter what the actual itinerary is, ie whether the flights are direct or not. The model can easily be adapted to the situation where, for managerial or technical reasons, fare products of all flights serving a market must be identical. This is achieved through the following variable substitution:

$$t^1_p = t^1_{p,f} \qquad \forall f \in F^1,\ \forall p \in P(f),$$

which ensures that a single price is published for each fare product, without regard to the flight on which this product is bought. Obviously, this restricts the solution space and might lead to lower revenue. It is now possible to display the bilevel program:

$$\max_{T^1, v^1, v^2} \sum_{f \in F^1} \sum_{p \in P(f)} \sum_{g \in G} t^1_{p,f} v^1_{p,f,g}$$

$$\text{s.t.} \sum_{f \in F^1(k)} \sum_{p \in P(f)} \sum_{g \in G} v^1_{p,f,g} \leq \overline{\sigma}_k \quad \forall k \in K$$

$$\sum_{f \in F^1(k)} \sum_{p \in P(f)} \sum_{g \in G} v^1_{p,f,g} \geq \underline{\sigma}_k \quad \forall k \in K$$

$$\sum_{f \in F^1(k)} \sum_{p \in P(f)} \sum_{g \in G} v^1_{p,f,g}\, t^1_{p,f} \leq \overline{\rho}_k \quad \forall k \in K$$

$$\sum_{f \in F^1(k)} \sum_{p \in P(f)} \sum_{g \in G} v^1_{p,f,g}\, t^1_{p,f} \geq \underline{\rho}_k \quad \forall k \in K$$

$$\min_{v^1, v^2} \sum_{f \in F^1} \sum_{p \in P(f)} \sum_{g \in G} c^1_{p,f,g}(\alpha, \beta) v^1_{p,f,g} + \sum_{f \in F^2} \sum_{p \in P(f)} \sum_{g \in G} c^2_{p,f,g}(\alpha, \beta) v^2_{p,f,g}$$

$$\text{s.t.} \sum_{\substack{p \in P(f) | \\ b(p)=b}} \sum_{g \in G} v^1_{p,f,g} \leq l_{b,f} \quad \forall b \in B(f),\ \forall f \in F^1$$

$$\sum_{f\in F^1(k)}\sum_{p\in P(f)} v^1_{p,f,g} + \sum_{f\in F^2(k)}\sum_{p\in P(f)} v^2_{p,f,g} = d_k h_{g,k} \quad \forall g \in G, \forall k \in K$$

$$\sum_{f|s\in S(f)}\sum_{p\in P(f)}\sum_{g\in G} v^1_{p,f,g} \leq u_s \forall s \in S$$

where

$$c^i_{p,f,g}(\alpha,\beta) = t^i_{p,f} + \alpha_{g,k} \times DUR_f + \beta_{g,k} \times QOS_{p,f}$$

for i $\in\backslash\{1,2\}$.

Illustrative example

Consider a two-market network (markets A–C and A–D) served by the leader airline and by two flights from competitors (see Figure 1). The leader operates its flights through a hub and uses three aircraft, one per leg. Leg capacities are equal to 200, 130 and 110 for legs 'a', 'b' and 'c', respectively.

Assuming a single fare product, each flight is characterised by two attributes: fare and 'disutility', which are displayed in Table 1.

Demand is split into two groups endowed with their own disutility factor, that is, their monetary perception of one unit of disutility. These values may vary between markets (see Table 2).

Dropping the fare product index p for clarity, the mathematical formulation of this instance takes the form

$$\max_{t,v} \quad t^1_1 v^1_{1,1} + t^1_1 v^1_{1,2} + t^1_2 v^1_{2,1} + t^1_2 v^1_{2,2}$$

$$\begin{aligned}\min_{v} \quad & [t^1_1 + (5\times 50)]v^1_{1,1} + [1{,}000 + (5\times 90)]v^2_{1,1} \\ & + [t^1_1 + (1\times 50)]v^1_{1,2} + [1{,}000 + (1\times 90)]v^2_{1,2} \\ & + [t^1_2 + (8\times 60)]v^1_{2,1} + [850 + (8\times 80)]v^2_{2,1} \\ & + [t^1_2 + (1\times 60)]v^1_{2,2} + [850 + (1\times 80)]v^2_{2,2}\end{aligned}$$

$$\begin{aligned}\text{s.t.} \quad & v^1_{1,1} + v^1_{1,2} + v^1_{2,1} + v^1_{2,2} \leq 200 \\ & v^1_{1,1} + v^1_{1,2} \leq 130 \\ & v^1_{2,1} + v^1_{2,2} \leq 110 \\ & v^1_{1,1} + v^2_{1,1} = 100 \\ & v^1_{1,2} + v^2_{1,2} = 450 \\ & v^1_{2,1} + v^2_{2,1} = 60 \\ & v^1_{2,2} + v^2_{2,2} = 385 \\ & v^1_{1,1}, v^1_{1,2}, v^1_{2,1}, v^1_{2,2}, v^2_{1,1}, v^2_{1,2}, v^2_{2,1}, v^2_{2,2} \geq 0\end{aligned}$$

Figure 1: Example network

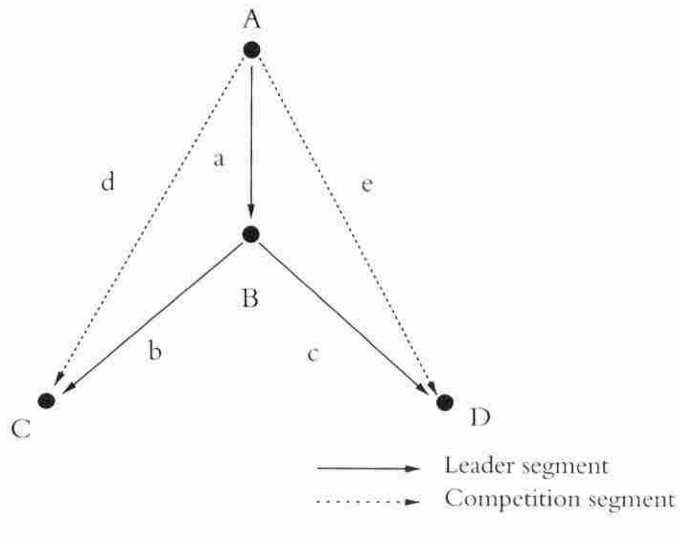

Table 1

Flight	*Market*	*Segments*	*Fare*	*Disutility*
1 (Leader)	A–C	a,b	t_1^1	50
2 (Leader)	A–D	a,c	t_2^1	60
1 (Competition)	A–C	d	\$1000	90
2 (Competition)	A–D	e	\$850	80

Table 2

Group	*Market*	*Demand*	*Disutility factor (\$)*
1	A–C	100	5
1	A–D	60	8
2	A–C	450	1
2	A–D	385	1

On this simple example, the model aims for the optimal mix of passengers, given the capacity of the bottleneck leg 'a'. The optimum fares are t_1^1 =\$1,200 and t_1^2 =\$870. One the one hand, these values make the leader's A-C flight attractive to group 1 users, since their perceived cost of \$1,450 [\$1,200 + (\$5 × 50)] is less or equal to that of the competition [\$1,000 + (\$5 × 90)]. On the other hand, users of the second group will fly the competition between A and C. On market A–D, the fare is attractive to both groups. Owing to capacity constraints, only 40 users of the second group are accommodated by the leader airline.

This solution achieves the best use of leg a's limited capacity, with a total revenue of \$207,000. This is an increase of 9.2 per cent over the strategy consisting in matching the competition's fares. The increase is equal to 5.6 per cent over a market-based strategy that optimises sequentially over markets A–C and A–D, in that order, and to 1.7 per cent when the sequence is reversed.

Solution methods

Although this paper focuses on a model, algorithmic issues are briefly discussed. Currently, algorithms for tackling bilevel models of the size proposed in this paper simply do not exist. One possible strategy would be to replace the lower-level problem by its primal–dual (KKT) optimality conditions, which are necessary and sufficient. This yields a single-level nonlinear program of greater dimensionality than the original problem, and involves hard complementarity constraints. This primal–dual problem can be reformulated as a mixed linear integer program and solved by a commercial code. Since this approach is only practical for small problem instances, the ideas presented in Labbé *et al.* (1998, 1999) or Brotcorne *et al.* (2000, 2001) were adopted. In particular, an inverse optimisation procedure was considered whereby an optimal fare schedule corresponding to a given lower level solution can easily be recovered. Heuristics were also designed to generate 'good' lower level solutions, and rules based on the problem's structure that allow the restriction of the search space were implemented. This enabled medium-sized instances to be solved for a near (1 per cent) optimum within minutes of CPU time.

Parameter calibration

It goes without saying that careful calibration of the behavioural parameters is essential to ensure the quality and the reliability of the fares and capacity allocation schemes generated by the model. The procedure developed to carry out this task is based on inverse optimisation, a well-known methodology in the physical sciences. A detailed exposition of this method is beyond the scope of this paper, but its basic ideas are easy to grasp. One can, with sufficiently accurate historical sales records, observe buying patterns and prices paid by the customers. By making the assumption that this buying behaviour actually depends on the parameters one wishes to calibrate, one may run the mathematical model 'backwards', the input being the historical records, and the output the model parameters. In particular, one seeks parameter values for which the model, in its 'forward' mode, generates an output that matches historical prices as closely as possible. The challenging aspects of this procedure lie in the formulation of the backward model and its numerical solution. It is remarkable that the problem induced by the backward formulation possesses a bilevel structure similar to that of the 'forward' model, and can therefore be addressed by similar algorithmic techniques.

PRACTICAL ISSUES

Data availability

This model has been implemented and validated in close collaboration with a major North American carrier. It relies on information that is generally available from information systems used by airlines. One must not lose sight, however, of the fact that seat inventory control and fare optimisation are data-intensive processes. Significant amounts of data are required from the model, including flight schedules, demand forecasts, fare description, etc. The parameter estimation procedure sketched above also requires historical sales data of sufficient detail and breadth. Major airlines treat and archive these data using complex information systems which, unfortunately, rarely use common data representation. A careful interpretation of the data retrieved from the various databases is a prerequisite to the construction of a coherent data model.

Dynamics

The model is inherently static and does not take into account the dynamic aspects of the airline fare optimisation problem; it takes a static snapshot of the current situation and suggests prices which are best for the leader, given the actual fares offered by the competition and the behaviour of the passengers. On the North American trans-border market, for instance, major carriers subscribe to a fare filing system known as ATPCo. By publishing their fares, participating airlines can access, in real time, information about the fares and products proposed by their competitors. This information is updated every four hours, so that fare optimisation must be performed several times a day to take advantage of the latest published fares of the competitors.

Thus, the model should run every time new fares are published. For each run, demand estimates and available capacity must be updated so that the input data reflect the latest bookings or any modification to the schedule and/or available capacity. Needless to say, solution algorithms must be efficient enough to solve real-life instances within minutes of computing time. In that perspective, it is appropriate to describe fare optimisation, at least on highly competitive markets, as a semi real-time process.

Randomness

A key assumption underlying this approach is that it leaves aside demand randomness. Or does it? Randomness may be confused with heterogeneity of demand, which is fully considered in the model, and which accounts for a large part of demand variability. In the mid-term, demand estimates are believed to be fairly accurate. Of course, demand can be stimulated by ticket sales or other marketing practices, but this should not be the prime focus of a strategic pricing model. Hence the trade-off achieved in favour of full network accounting versus the short-term, day-to-day variability of demand. On the pragmatic side, the extension of this approach to a stochastic environment would require the estimation of demand elasticities over the entire network, a hazardous task at best.

CONCLUSION

The aim of this paper was to propose a new modelling approach to pricing and fare optimisation in the airline industry. The use of bilevel programming allows the maximisation of revenue while taking into account in a detailed fashion the behaviour of the passengers, as well as the complex topology of airline networks. This is in sharp contrast to models based on econometric demand estimates, or stochastic models that focus on a single leg or a single market. It is believed that what is included in the model more than offsets the features left out, ie the dynamics and variability of demand. In particular, it deals with the simultaneous optimisation of the two components of the revenue management process (pricing and seat allocation) and could even address demand forecasting in a crude fashion by incorporating a dummy 'no-travel' alternative to passengers.

Of course, there is a price to pay for the features provided by the model. In particular, the booking limits of the competing airlines are not known, although estimates, as well as global capacities, were they available, could easily be incorporated into the model's constraints. Also, as mentioned earlier, demand variability is not addressed directly, although its impact would be mitigated by running the model on a frequent basis, using updated demand estimates.

Notes

1 The fare basis code is a concise notational device used to label a published fare product and describe its main attributes. There is no standard rule for composing fare basis codes, but airlines usually follow generally agreed conventions. For instance, the first letter of a fare basis code denotes the booking class to which the corresponding fare product belongs. Other characters and digits may represent advance booking restrictions, Saturday-night stay-over, refundability, etc.

2 Note that, in practice such a function mapping fare products to booking classes would be trivial, since the class to which a product belongs is given by the first letter of its 'fare basis code'.

Acknowledgment

This paper was supported by NSERC, NATEQ and MITACS.

References

Belobaba, P. P. (1987a) 'Airline yield management: An overview of seat inventory control', *Transportation Science*, **21**, 2, 63–73.

Belobaba, P. P. (1987b) 'Application of a probabilistic decision model to airline seat inventory control', *Operations Research*, **37**, 2, 183–197.

Belobaba, P. P. (1998) 'The evolution of airline yield management: Fare class to origin–destination inventory control', in Butler, G. F. and Keller, M. R. (eds) *Handbook of Airline Marketing*, Aviation Week Group, McGraw-Hill, 285–302.

Brotcorne, L., Labbé, M., Marcotte, P. and Savard, G. (2000) 'A bilevel model and solution algorithm for a freight tariff setting problem', *Transportation Science*, **34**, 317–324.

Brotcorne, L., Labbé, M., Marcotte, P. and Savard, G. (2001) 'A bilevel model for toll optimization on a multicommodity transportation network', *Transportation Science*, **35**, 1–14.

Brumelle, S. L. and McGill, J. I. (1993) 'Airline seat allocation with multiple fare classes', *Operations Research*, **41**, 1, 127–137.

Brumelle, S. L., McGill, J. I., Oum, T. H., Sawaki, K. and Tretheway, M. W. (1990) 'Allocation of airline seats between stochastically dependent demands', *Transportation Science*, **24**, 3, 183–192.

Curry, R. (1990) 'Optimal airline seat allocation with fare classes nested by origins and destinations', *Transportation Science*, **24**, 3, 193–204.

De Boer, S. V., Freling, R. and Piersma, N. (1999) 'Stochastic programming for multiple-leg network revenue management', Tech. Rep. EI-9935/A, ORTEC Consultants, Gouda, The Netherlands.

Feng, Y. and Xiao, B. (2001) 'A dynamic seat inventory control model and its optimal policy', *Operations Research*, **49**, 6, 938–949.

Gallego, G. and Ryzin, G. V. (1994) 'Optimal dynamic pricing of inventories with stochastic demand over finite horizons', *Management Science*, **40**, 8, 999–1020.

Gallego, G. and Ryzin, G. V. (1997) 'A multiproduct dynamic pricing problem and its applications to network yield management', *Operations Research*, **45**, 1, 24–41.

Garvett, D. and Michaels, L. (1998) 'Price parraying: A direction for quick, decisive and profit-maximizing pricing', in Butler, G. F. and Keller, M. R. (eds) *Handbook of Airline Marketing*, Aviation Week Group, McGraw-Hill, 333–348.

Glover, F., Glover, R., Lorenzo, J. and McMillan, C. (1992) 'The passenger mix problem in the scheduled airlines', *Interfaces*, **12**, 3, 73–79.

Kimes, S. (1990) 'Yield management: A tool for capacity-constrained service firms', *Journal of Operations Management*, **8**, 348–363.

Kretsch, S. S. (1998) 'Airline fare management and policy', in Jenkins, D. (ed.) *Handbook of Airline Economies*, Aviation Week Group, McGraw-Hill, 477–482.

Kuyumcu, A. and Garcia-Diaz, A. (2000) 'A polyhedral graphs theory approach to revenue management in the airline industry', *Computers and Industrial Engineering*, **38**, 3, 375–395.

Labbé, M., Marcotte, P. and Savard, G. (1998) 'A bilevel model of taxation and its application to optimal highway pricing', *Management Science*, **44**, 1595–1607.

Labbé, M., Marcotte, P. and Savard, G. (1999) 'On a class of bilevel programs', in Pillo, G. D. and Giannessi, F. (eds) *Nonlinear Optimization and Related Topics*, Kluwer, Dordrecht, 183–206.

Lautenbacher, C. and Shaler, S. (1999) 'The underlying Markov decision process in the single-leg airline yield-management problem', *Transportation Science*, **33**, 2, 136–146.

Lee, T. C. and Hersh, M. (1993) 'A model for dynamic airline seat inventory control withmultiple seat bookings', *Transportation Science*, **27**, 3, 252–265.

Littlewood, K. (1972) 'Forecasting and control of passenger bookings', *AGIFORS Symposium Proceedings*, **12**, 95–117.

McGill, J. and Van Ryzin, G. (1999) 'Revenue management: Research overview and prospects', *Transportation Science*, **33**, 2, 233–256.

Pak, K. and Piersma, N. (2002) 'Airline revenue management: An overview of OR techniques 1982–2002', Econometric Institute, Erasmus University, Rotterdam, The Netherlands.

Robinson, L. W. (1993) 'Optimal and approximate control policies for airline booking with sequential nonmonotonic classes', *Operations Research*, **43**, 2, 252–263.

Smith, B., Leimkuhler, J. and Darrow, R. (1992) 'Yield management at American Airlines', *Interfaces*, **22**, 1, 8–31.

Subramanian, J., Shaler, S., and Lautenbacher, C. (1999) 'Airline yield management with overbooking, cancellations, and no-shows', *Transportation Science*, **33**, 2, 147–167.

Talluri, K. and van Ryzin, G. (1998) 'An analysis of bid-price controls for nework revenue management', *Management Science*, **44**, 11, 1577–1593.

Talluri, K. T. (2001) 'Airline revenue management with passenger routing control: A new model with solution approaches', *International Journal of Services Technology and Management*, **2**, 1/2, 102–115.

Vinod, B. (1998) 'Origin-and-destination yield management', in Jenkins, D. (ed.) *Handbook of Airline Economies*, Aviation Week Group, McGraw-Hill, 459–468.

Weatherford, L. and Bodily, S. (1992) 'A taxonomy and research overview of perishable-asset revenue management: Yield management, overbooking, and pricing', *Operations Research*, **40**, 5, 831–844.

Wollmer, R. D. (1992) 'An airline seat management model for a single leg route when lower fare classes book first', *Operations Research*, **40**, 1, 26–37.

Wong, J., Koppelman, F. and Daskin, S. (1993) 'Flexible assignment approach to itinerary seat allocation', *Transportation Research B*, **27B**, 1, 33–48.

You, P.-S. (1999) 'Dynamic pricing in airline seat management for flights with multiple flight legs', *Transportation Science*, **33**, 2, 192–206.