

DATA-DRIVEN JOINT OPTIMIZATION OF PRICING AND SEAT ALLOCATION IN TRAINS

NAOTO KONNO¹ & ARVIND U. RAGHUNATHAN²

¹Mitsubishi Electric Corp, Japan

²Mitsubishi Electric Research Laboratories Inc., USA

ABSTRACT

Revenue management (RM), a management science method, employs demand predictions to maximize revenue. Since its introduction in the airline industry in the 1980s, it is now widely used in various industries such as hotel, retail, and railway, among others. Compared to the airline industry, penetration has been slow in the railway industry. In recent years, RM has seen increased adoption in the railway industry. Existing approaches to RM in trains have mostly considered the determination of price and capacity separately. In this paper, we consider the problem of maximizing revenue by optimizing over price and capacity simultaneously. We make three important contributions to this problem. For the first time, we consider the problem of train sizing, i.e. number of coaches in each fare class. Second, instead of logit-based models estimated from price-demand data, we consider data-driven models that alleviate the problem of modelling errors. Finally, we formulate the optimization as an integer linear program as opposed to the nonlinear formulations that result from logit-based models. Thus, we present data-driven revenue maximization to jointly determine pricing, seat allocation, train sizing for a single fare class over multiple legs and train lines. Our simulation results targeting Nozomi on the Tokaido-Sanyo Shinkansen in Japan show that it is possible to formulate a railway operation plan to maximize the operation profit by applying our proposed method. In our numerical experiments, the proposed approach can identify solutions that increase profit per day by 50%, with a computation time of less than 1 second.

Keywords: data-driven, optimization, pricing, capacity allocation, train sizing, revenue management, railways, integer linear program.

1 INTRODUCTION

Revenue management (RM) (or yield management) is a fundamental problem in service operations such as airlines, bus companies, hotel industry, manufacturing companies, and railways among others. RM studies the problem of allocating a firm's resources at different rates to maximize revenues. Seat inventory control, a key concept in RM, is concerned with the problem of determining whether to accept a request at a certain fare under the condition that its value is lost after the sales season. The fares vary based on the class (first class or economy class) and the date of purchase. For example, the fare is more expensive when purchased closer to the date of travel.

RM has been widely employed in the context of airline operations. Seat inventory control has been investigated in Glover et al. [1], Belobaba [2], Davis [3], Weatherford and Bodily [4], McGill and van Ryzin [5] and Talluri and van Ryzin [6]. In the airlines industry, a leg refers to a single take-off and landing. Hence, seat booking for passengers with one or more stops from origin to destination it is necessary to consider the seat allocation by considering the airline's entire network operations. In the context of airlines, the single leg and multi-leg inventory control have been considered.

Train line refers to a train running between terminal stations that stops at multiple intervening stations. A leg (or a segment) in rail operations refers to journey between successive stations at which the train makes a stop. Train lines are inherently multi-leg operations. Hence, the seat allocation for all possible origin-destinations (stations on the train line) has to be considered simultaneously. The application of RM in freight railways can be



traced to the work of Strasser [7]. To the best of our knowledge, Ciancimino et al. [8] was the first work on RM for passenger railways. Ciancimino et al. [8] considered the problem of determining seat allocations on a single fare class, multiple legs of the train line to maximize the revenue. A deterministic linear programming formulation was proposed for the case of deterministic demand and a nonlinear programming formulation was proposed to handle demand uncertainty. The authors ignored the integrality of the seat allocation decisions. You [9] formulated the seat allocation among two fare-classes, multiple legs on a train line as a mixed integer nonlinear program and proposed a heuristic solution approach. Armstrong and Meissner [10] present an overview of RM approaches and models in the context of railways. An extension of the two-fare class model of You [9] to the multi-fare class was presented in [10]. Hetrakul and Cirillo [11] developed a nonlinear programming formulation that jointly optimized the prices and seat allocation based on a latent choice model. A multinomial logit model was used to predict the passenger choices. Wang et al. [12] presented a two-stage stochastic formulation incorporating passenger choice for determining optimal seat allocations. The resulting policies were compared in simulations. Zhao and Zhao [13] considered the problem of optimal seat allocation on multiple train lines under customer choice model. They solved the problem using a heuristic algorithm and demonstrated their results on Beijing–Shanghai High Speed Rail (HSR). Zhai et al. [14] optimized seat allocation for multiple lines serving the same set of origin and destinations in order to maximize the passenger throughput. This study does not directly pertain to RM.

The survey of the literature on RM in railways reveals that:

1. Train sizing, in terms of number of coaches in each fare class, has not been considered as a decision variable. This aspect is particular to railways and not airlines. For instance, the operator can choose to vary the number of first- and second-class coaches for a particular train line based on the demand.
2. Logit-based models are primarily used for modelling price sensitivity when optimizing jointly for prices and seat allocations. However, such models are estimated from data and can introduce errors in the results of the optimization. Further, the use of such models introduces nonlinearity in the problem and makes them intractable.
3. Heuristic algorithms such as particle swarm optimization, simulated annealing are the dominant solution approaches. These algorithms claim to be able to find good feasible solutions. The algorithms do not provide an estimate of the optimality gap making it difficult to assess the quality of the solution. Hence, practitioners resort to running these algorithms for a long time in the hope of finding better solutions.

In this paper, we consider the data-driven joint optimization of pricing, seat allocation, sizing to maximize revenue for a single fare class over multiple legs and train lines. In particular, we: (i) consider number of coaches as a decision variable; (ii) assume that price-demand function is provided as pairs of price and demand which are derived from user surveys. In particular, we do not estimate a logit-based model from this data; (iii) present an exact approach to solve the problem based on an integer linear programming (ILP) formulation. Note that the data-driven approach allows us to model the problem as an integer linear optimization instead of a nonlinear optimization problem. Further, the formulation ensures that the variables modelling seat allocations and number of coaches are integer valued which has been ignored in prior works, especially when the formulation becomes nonlinear. ILP formulations can be easily solved using solvers such as CPLEX Optimizer [15], Gurobi [16], and FICO Xpress Solver [17]. ILP solvers are robust and are guaranteed to find the optimum. Further, the solvers easily scale to the problem sizes of interest. The results in Section 4 demonstrate this.



2 NOTATION

The set of integers and non-negative integers are denoted by \mathbb{Z} and \mathbb{Z}_+ respectively. For a positive integer m , $[m]$ and $[m]_0$ denote respectively the sets $\{1, \dots, m\}$ and $\{0, \dots, m\}$. Let \mathcal{T} be the set of train lines, n_t be the number of trains operated on the line over a day and $S(t)$ be the set of stops that are served by the train line $t \in \mathcal{T}$. Let $S = \cup_{t \in \mathcal{T}} S(t)$ be the set of all stops that are served by the train lines. Further we assume that the set of all stations served by the train lines can be ordered so $S := [|S|]$ i.e. if $i, j \in S(t)$ with $i < j$ then train line t stops at i before j for all $t \in \mathcal{T} : i, j \in S(t)$. The minimum and maximum number of coaches that can be attached to the train line $t \in \mathcal{T}$ are \underline{u}_t and \bar{u}_t respectively. Denote by C^{max} the number of seats in a coach of the train. Let $c_{t,l}$ be the cost of operating l number of coaches on train line $t \in \mathcal{T}$. A trip is denoted as a pair of stations $(i, j) \in S(t) \times S(t)$ such that a passenger can travel from station $i \in S(t)$ to station $j \in S(t)$ on the train line t . The set $J(t) \subset S(t) \times S(t)$ denotes all possible trips on the train line $t \in \mathcal{T}$ i.e. $J(t) := \{(i, j) \mid i, j \in S(t), i < j\}$. Let, $\underline{x}_{t,ij}(t), \bar{x}_{t,ij}(t) \in \mathbb{Z}$ denote the minimum and maximum number of seats for the trip $(i, j) \in J(t)$ on the train line $t \in \mathcal{T}$. Let $J := \cup_{t \in \mathcal{T}} J(t)$ be the set of all trips that can be taken using any of the train lines. For each trip $(i, j) \in J$ there exists a set $PD(i, j) := \{(p_{ij,1}, d_{ij,1}), \dots, (p_{ij,K(i,j)}, d_{ij,K(i,j)})\}$, representing the price-daily demand pairs for the trip (i, j) with $p_{ij,k} > 0, d_{ij,k} \in \mathbb{Z}_+$ for $k \in [K(i, j)]$. Let $J(t, q) \subset J(t)$ for $q \in S(t)$ be defined as $J(t, q) := \{(i, j) \mid i \leq q < j, (i, j) \in J(t)\}$. The set $J(t, q)$ defines the set of trips that cover the leg of the train line t from station q . By definition, $J(t, q) = \emptyset$ if q is the last station on the train line t .

3 INTEGER LINEAR PROGRAMMING FORMULATION

We present an ILP formulation for the joint optimization of sizing, price and seat allocation for a single fare class, multi-leg RM over multiple train lines. The variables in the problem are

- $y_{t,l} \in \{0,1\}$ denote the choice of l coaches for $l = \underline{u}_t, \dots, \bar{u}_t$ on train line $t \in \mathcal{T}$;
- $x_{t,ij} \in \mathbb{Z}_+$ denote the number of seats that are allocated for the trip $(i, j) \in J(t)$ on train line $t \in \mathcal{T}$;
- $s_{ijk} \in \mathbb{Z}_+$ denote the number of seats that are sold at the price $p_{ij,k}$ on the trip $(i, j) \in J, k \in [K(i, j)]$;
- $z_{ijk} \in \{0,1\}$ denote the choice of price $p_{ij,k}$ for the segment $(i, j) \in J, k \in [K(i, j)]$. Note that the realized demand at price $p_{ij,k}$ is $d_{ij,k}$ with $(p_{ij,k}, d_{ij,k}) \in PD(i, j)$.

The optimization problem for maximizing profit is

$$\max \sum_{(i,j) \in J} \sum_{k \in [K(i,j)]} p_{ij,k} s_{ijk} - \sum_{t \in \mathcal{T}} n_t \cdot \left(\sum_{l=\underline{u}_t}^{\bar{u}_t} c_{t,l} y_{t,l} \right), \quad (1)$$

$$\text{s.t. } \sum_{(i,j) \in J(t,q)} x_{t,ij} \leq C^{max} \cdot \sum_{l=\underline{u}_t}^{\bar{u}_t} l \cdot y_{t,l} \quad \forall q \in S(t), t \in \mathcal{T}, \quad (2)$$

$$\underline{x}_{t,ij} \leq x_{t,ij} \leq \bar{x}_{t,ij} \quad \forall (i, j) \in J(t), t \in \mathcal{T}, \quad (3)$$

$$\sum_{k \in [K(i,j)]} s_{ijk} \leq \sum_{t \in \mathcal{T} : (i,j) \in J(t)} n_t \cdot x_{t,ij} \quad \forall (i, j) \in J, \quad (4)$$

$$s_{ijk} \leq d_{ij,k} z_{ijk} \quad \forall (i, j) \in J, k \in [K(i, j)], \quad (5)$$

$$\sum_{k \in [K(i,j)]} z_{ijk} = 1 \quad \forall (i, j) \in J, \quad (6)$$

$$\sum_{l=\underline{u}_t}^{\bar{u}_t} y_{t,l} = 1 \quad \forall t \in \mathcal{T}, \quad (7)$$

$$y_{t,l} \in \{0,1\} \quad \forall l = \underline{u}, \dots, \bar{u}, t \in \mathcal{T}, \quad (8)$$

$$s_{ijk} \in \mathbb{Z}_+ \quad \forall (i,j) \in J, k \in [K_{ij}], \quad (9)$$

$$x_{t,ij} \in \mathbb{Z}_+ \quad \forall (i,j) \in J(t), t \in \mathcal{T}, \quad (10)$$

$$z_{ijk} \in \{0,1\} \quad \forall (i,j) \in J, k \in [K_{ij}]. \quad (11)$$

The first term in the objective function models the revenue that is obtained on each segment as the product of the price $p_{ij,k}$ and the number of seats that are sold s_{ijk} . The second term in the objective models the cost incurred in operating a certain number of coaches and introduces a trade-off between revenue and operating costs. Note that the formulation does not impose any assumptions on the functional form of c_l . Constraint in eqn (2) ensures that the total number of seats allocated to all of the trips in $J(t, q)$ that include the segment starting from $q \in S(t)$ of the train line $t \in \mathcal{T}$ does not exceed the total capacity of the train. Constraint in eqn (3) requires the number of seats allocated for trip $(i, j) \in J(t)$ on train line $t \in \mathcal{T}$ to be between the specified lower ($x_{t,ij}$) and upper ($\bar{x}_{t,ij}$) limits. Constraint in eqn (4) limits the number of seats sold $\sum_{k \in [K_{ij}]} s_{ijk}$ for trip $(i, j) \in J$ to be upper bounded by the total seat allocation over all train lines serving that trip. Constraint in eqn (5) limits the number of seats sold s_{ijk} for trip $(i, j) \in J$ at price $p_{ij,k}$ by the demand $d_{ij,k}$. The equality in eqn (6) enforces that only one of the available price alternatives is chosen for each of the trips $(i, j) \in J$. Eqn (7) ensures that only one of the binary variables $y_{t,l}$ takes a value of 1. The remaining constraints enforce the appropriate integrality of the variables. Note that the formulation does not provide the number of seats that are sold on each train line. However, this can be obtained by apportioning the number of seats sold ($\sum_{k \in [K_{ij}]} s_{ijk}$) to each train line such that the assignment does not exceed the available seats for the trip ($n_t \cdot x_{t,ij}$) on the train line.

The constraints in the formulation are all linear and the decision variables are binary or integer valued. Hence, our formulation belongs to the class of ILPs. A number of commercial software packages have been developed for the solution of ILPs, namely CPLEX Optimizer [15], Gurobi [16], and FICO Xpress Solver [17].

We conclude the section by noting the following. Suppose a logit model is available for price-demand elasticity. Then a set with pairs of price and demand as in $PD(i, j)$ can be obtained by discretizing the logit-curve. An ILP formulation, so obtained, can alleviate the nonlinearity that is inherent when using the logit model. The number of price discretizations considered will influence the optimal solution. We will study the effect of the number of discretization points in Section 4.4.

4 NUMERICAL EXAMPLE

In order to showcase the effectiveness of our research, this section introduces an application example targeting high-speed railways in Japan. High-speed railways in Japan are called Shinkansen, and is a major means of transportation in Japan. In our study, we focus on “Tokaido-Sanyo Shinkansen” which is the one of the most famous Shinkansen lines in Japan. There are five train types “Nozomi”, “Hikari”, “Kodama”, “Sakura” and “Mizuho” in Tokaido-Sanyo Shinkansen. We use data from a variety of sources to estimate the demand data on Nozomi, which is the fastest type in Tokaido-Sanyo Shinkansen and used by a great deal of people. In the following section, we describe our approach to deriving the data required for the formulation. Section 4.1 presents our criteria for selecting the stations that are included in the study. This limits the set of trips that we consider in the formulation. Section 4.2 describes our method for estimating the number of passengers who take Nozomi from each station over a day. Section 4.3 provides the setting of the parameters used in the

optimization formulation of Section 3. Section 4.4 describes the results from solving the optimization formulation in Section 3 under different parameter settings.

4.1 Selection of criteria for target station

Nozomi is a type of Shinkansen that operates between Tokyo and Hakata, and Table 1 lists the stations at which it stops. There are 15 Nozomi stops in total, but Himeji, Fukuyama, Tokuyama, and Shin-Yamaguchi are excluded from the study as they are only served by a few services. In this research, when there are two stations in one prefecture, only one of them is selected as representative in order to estimate the demand by using the prefecture's mutual departure and arrival table [18]. For instance, Tokyo Station and Shinagawa Station are in Tokyo prefecture, while Kokura and Hakata are in Fukuoka prefecture. Therefore, Tokyo and Hakata Station are respectively used as a representative of Tokyo and Fukuoka prefectures. This is done primarily due to a lack of disaggregated ridership data by stations in a prefecture. The column titled "Target station" in Table 1 shows the stations targeted in this study.

Table 1: List of stations.

Station name	Nozomi stop station	Target station
Tokyo	✓	✓
Shinagawa ²	✓	
Shin-Yokohama	✓	✓
Nagoya	✓	✓
Kyoto	✓	✓
Shin-Osaka	✓	✓
Shin-Kobe	✓	✓
Himeji	✓ ¹	
Okayama	✓	✓
Fukuyama	✓ ¹	
Hiroshima	✓	✓
Tokuyama	✓ ¹	
Shin-Yamaguchi	✓ ¹	
Kokura ²	✓	
Hakata	✓	✓

¹Only some trains stop.

²Not representative stations in each prefecture.

4.2 Estimating Shinkansen ridership

In this section, we describe our method for estimating the passenger traffic information. We describe the method for estimating the number of passengers boarding and deboarding the Shinkansen at each main station in a day in Section 4.2.1. Section 4.2.2 describes the method for estimating the flow of passengers between stations.

4.2.1 Number of passengers at each stations per day

West Japan Marketing Communications Inc. records the average number of passengers boarding Shinkansen every day from each main station [19]. We assume that the ratio of

Shinkansen passengers to other railway passengers is 3:7 if some stations' data also includes non-Shinkansen passengers. The second column in Table 2 shows the number of one way Shinkansen users of our target stations by using this data.

4.2.2 Number of passengers on different trips

In Section 4.2.1 we estimated the number of passengers riding the train from each station. However, we have to estimate the number of passengers moving between stations in order to obtain the demand for each trip. The Ministry of Land, Infrastructure, Transport and Tourism surveyed the number of passengers moving between prefectures by Shinkansen or limited express train [18]. We utilize this information to estimate passengers flow.

As shown in Table 1, we consider the stations Tokyo, Shin-Yokohama, Nagoya, Kyoto, Shin-Osaka, Shin-Kobe, Okayama, Hiroshima and Hakata. In addition, the prefectures in which these stations belong to are Tokyo, Kanagawa, Aichi, Kyoto, Osaka, Hyogo, Okayama, Hiroshima, and Fukuoka in that order. These prefectures are also our target prefectures in this study. We used "One-way / round-trip inter-prefecture inter-arrival table" made available by the Ministry of Land, Infrastructure, Transport and Tourism [18] to obtain the percentage flows of passengers listed in fourth column of Table 2. The fifth column is the result multiplying the number of Shinkansen passengers from each station (column 2) with flow percentages (fourth column). The fifth column provides the number daily Shinkansen passengers between different stations.

4.2.3 Estimation for Nozomi passengers

In this section, we estimate the number of passengers using the Nozomi train (sixth column in Table 2) from the total number of Shinkansen passengers (fifth column). There are three types of Shinkansen, Nozomi, Hikari and Kodama from Tokyo. In addition, there are five types of Shinkansen, Nozomi, Hikari, Kodama, Sakura and Mizuho from Shin-Osaka. In order to estimate the number of Nozomi passengers, we consider the proportion of Nozomi departing from Tokyo to Hakata and from Shin-Osaka to Hakata. According to the JR Outing Net [20], about 62% of the Shinkansen from Tokyo is of Nozomi type, and about 50% of the Shinkansen from Shin-Osaka is of Nozomi type. Assuming that the number of passengers in Nozomi follows this ratio, we can estimate Nozomi passengers in sixth column of Table 2. The Nozomi consists of three different fare classes – business, reserved and non-reserved. We only focus on one fare class, reserved seat class. Assuming that 66% of Nozomi's passengers is reserved seat passengers, we can estimate reserved seat passengers in seventh column of Table 2.

4.3 Problem parameters

In this section we describe the values for the different parameters in the formulation provided in Section 3.

- **Train lines:** We consider 7 train lines ($\mathcal{T} = \{1, \dots, 7\}$) as specified in Table 3. The number of trains that run on each of the train lines (n_t) is provided in the fourth column of Table 3. The cost for operating a coach on each of the train lines is provided in the fifth column of Table 3. The cost of operating l coaches on line t ($c_{t,l}$) is assumed to be l times the cost listed in the fifth column of Table 3. The cost for operating a coach consists of: (i) Electric cost (18.75 YEN/kilometer), (ii) Maintenance cost (YEN 10,000 per day), and (iii) Cleaning fee as Nozomi's operating expenses (YEN 1000 per train). Thus, the operating cost per coach varies by the train line. The coach capacity (C^{max}) is assumed to be 87.

Table 2: Information on number of passengers on different trips.

Departure	Total passengers	Arrival	Flow percentage ³	Flow passengers ⁴	Nozomi passengers ⁴	Reserved seat passengers ⁴
Tokyo	132,900	Shin-Yokohama	2.24	2,981	1,848	1,220
		Nagoya	17.57	23,350	14,477	9,555
		Kyoto	10.12	13,443	8,335	5,501
		Shin-Osaka	21.97	29,193	18,100	11,946
		Shin-Kobe	10.14	13,471	8,352	5,512
		Okayama	3.71	4,928	3,055	2,017
		Hiroshima	6.49	8,622	5,346	3,528
Shin-Yokohama	32,700	Hakata	1.98	2,634	1,633	1,078
		Nagoya	13.16	4,305	2,669	1,762
		Kyoto	11.59	3,789	2,349	1,551
		Shin-Osaka	18.39	6,013	3,728	2,460
		Shin-Kobe	9.00	2,943	1,825	1,204
		Okayama	3.33	1,090	676	446
		Hiroshima	6.04	1,976	1,225	809
Nagoya	62,670	Hakata	2.11	689	427	282
		Kyoto	4.23	2,650	1,643	1,084
		Shin-Osaka	12.76	7,998	4,959	3,273
		Shin-Kobe	4.46	2,792	1,731	1,143
		Okayama	1.86	1,165	722	477
		Hiroshima	3.74	2,342	1,452	958
		Hakata	4.25	2,661	1,650	1,089

Table 2: Continued.

Departure	Total passengers	Arrival	Flow percentage ³	Flow passengers ⁴	Nozomi passengers ⁴	Reserved seat passengers ⁴
Kyoto	37,600	Shin-Osaka	5.56	2,089	1,295	855
		Shin-Kobe	5.07	1,905	1,181	779
		Okayama	1.87	704	437	288
		Hiroshima	2.39	897	556	367
Shin-Osaka	79,300	Hakata	5.22	1,964	1,218	804
		Shin-Kobe	3.70	2,933	1,466	968
		Okayama	4.11	3,257	1,629	1,075
		Hiroshima	6.48	5,141	2,571	1,697
Shin-Kobe	9,300	Hakata	7.66	6,074	3,037	2,004
		Okayama	6.86	638	319	210
		Hiroshima	6.84	636	318	210
		Hakata	7.24	674	337	222
Okayama	20,160	Hiroshima	8.62	1739	869	574
		Hakata	2.98	601	300	198
Hiroshima	22,380	Hakata	9.86	2206	1,103	728

³Two decimal places display; ⁴Integer display.



- Set of stations: The set of stations under consideration are provided in the third column of Table 1. The ordering of the stations is the sequence in which the stations are visited. Table 3 specifies the origin and destination terminals for each train line. The train line is assumed to stop at all the intermediate stations. For instance, the Tokyo-Shin-Osaka train line (Route no 1 in Table 3) stops at Tokyo, Shin-Yokohama, Nagoya, Kyoto, Shin-Osaka, i.e. stops at all intermediate stations mentioned in the third column of Table 2. Using this sets $S(t), J(t), J(t, q)$ can all be inferred.
- Price-demand function: We use the demand data inferred in the seventh column of Table 2 and the current fares for these trips (provided in Table 4) to estimate a price-demand logit function. Logit price-response curve for all $(i, j) \in J$

$$d_{ij}(p_{ij}) = C_{ij} e^{-(a_{ij} + b_{ij} p_{ij})} / (1 + e^{-(a_{ij} + b_{ij} p_{ij})}). \quad (12)$$

The parameter b_{ij} is set to 0.001 for all $(i, j) \in J$. The demand (\hat{d}_{ij}) in the seventh column of Table 2 is observed at the price (\hat{p}_{ij}) stated in Table 4. We choose $a_{ij} = -b_{ij} \hat{p}_{ij}$ and hence, $C_{ij} = 2\hat{d}_{ij}$. We will discretize this logit curve at a set of prices to obtain our price-demand pairs. The price is discretized at intervals of Δp_{ij} as $\{\hat{p}_{ij} + k \cdot \Delta p_{ij}\}$ for $k = -n_{disc}, \dots, n_{disc}$ to obtain $K_{ij} = (2n_{disc} + 1)$ prices. The demand function in eqn (12) is evaluated at the discretized prices to yield $(2n_{disc} + 1)$ price-demand pairs. We employ this approach since we do not have access to the raw data. But this suffices to illustrate the computational efficiency of our approach. The values of $\Delta p_{ij}, n_{disc}$ are provided in the next section.

- Lower and upper limit on each trip: We do not impose the constraint in eqn (3). In other words, $\underline{x}_{t,ij} = 0$ and $\bar{x}_{t,ij} = \infty$ for all $t \in \mathcal{T}, (i, j) \in J(t)$. This is done due to the lack of data on this parameter.

4.4 Results

We will begin this section by first studying the effect on operating costs, revenue and profits from having number of coaches and price to be a decision variable in the optimization. In the rest of the section the price interval parameter (Δp_{ij}) is set to $(0.01 \cdot \hat{p}_{ij})$, i.e. 1% of the nominal price in Table 4. With this setting the price for each trip will be varied within $\pm n_{disc} \%$ of \hat{p}_{ij} . We will vary the number of price points (n_{disc}) in the following. We consider the following four different realizations of the optimization formulation described in Section 3:

- Baseline : In this realization, we fix the number of coaches on each train line to be 10 ($\underline{u}_t = \bar{u}_t = 10$). The parameter n_{disc} is set to 1. This implies that price for each trip is fixed to the value listed in Table 4. Hence, number of coaches and prices are not optimized. In this setting, the cost of operation is a fixed quantity since the number of coaches cannot be varied. The optimization algorithm finds a feasible allocation of seats according to the demand on each trip as specified in Table 2.
- FixedPrices: In this realization, we continue to use the parameter n_{disc} is set to 1. We allow the number of coaches on each train line to be a decision variable. The lower and upper limits for each train line is set as $\underline{u}_t = 6, \bar{u}_t = 16$. By varying the number of coaches in the optimization it is possible to reduce the operating costs and hence, increase profits. This realization allows to study the effect of variation in train sizes when prices are fixed.



Table 3: Nozomi's travel route.

Route no	Departure	Arrival	Trains per day	Cost per coach (YEN)
1	Tokyo	Shin-Osaka	56	20,656.25
2		Okayama	4	23675.00
3		Hiroshima	14	26,393.75
4		Hakata	33	31,043.75
5	Shin-Osaka	Okayama	4	14,000.00
6		Hiroshima	14	16,718.75
7		Hakata	37	21,368.75

Table 4: The fare table.

Departure	Arrival	Fare (YEN)
Tokyo	Shin-Yokohama	2,960
	Nagoya	11,090
	Kyoto	13,910
	Shin-Osaka	14,450
	Shin-Kobe	15,100
	Okayama	17,340
	Hiroshima	19,080
	Hakata	22,950
Shin-Yokohama	Nagoya	10,450
	Kyoto	13,250
	Shin-Osaka	14,120
	Shin-Kobe	14,770
	Okayama	17,010
	Hiroshima	18,760
	Hakata	22,630
Nagoya	Kyoto	5,800
	Shin-Osaka	6,560
	Shin-Kobe	8,180
	Okayama	11,290
	Hiroshima	14,230
	Hakata	18,540
Kyoto	Shin-Osaka	3,020
	Shin-Kobe	3,650
	Okayama	7,850
	Hiroshima	11,410
	Hakata	16,060
Shin-Osaka	Shin-Kobe	3,100
	Okayama	6,230
	Hiroshima	10,440
	Hakata	15,310
Shin-Kobe	Okayama	5,800
	Hiroshima	10,220
	Hakata	14,990
Okayama	Hiroshima	6,230
	Hakata	12,710
Hiroshima	Hakata	9,150

- FixedCoaches-5%: In this realization, we fix the number of coaches as specified in Baseline ($\underline{u}_t = \bar{u}_t = 10$). However, the price is allowed to vary on each trip. We use $n_{disc} = 5$ so the price for each trip is varied within $\pm 5\%$ of \hat{p}_{ij} to yield $K_{ij} = 11$ price-demand pairs for each trip.
- Optimized-5%: In this realization, the number of coaches on each train line and the price are decision variables in the optimization as proposed in this paper. The prices on each trip are allowed to vary within $\pm 5\%$ of \hat{p}_{ij} as in FixedCoaches-5%.

The resulting optimization problems for each of the four realizations are solved using Gurobi 9.0 [16]. In all instances, the solver took less than 0.1 second on a standard desktop to obtain the optimal solutions. Further, the solution is obtained at the root node of the tree. In all cases, the number of seats that are sold on each trip is equal to the demand on that trip. This is to be expected when the capacity that is available on each trip is in excess of the demand. An inspection of the solution for each of these cases also reveals that no seats are allocated to certain trips on some of the train lines. The number of coaches on each train line, total cost of operation, revenue, and profit is listed in Table 5. The number of coaches in Table 5 are listed in the order of the train lines specified in Table 3.

Table 5: Comparison of different realizations of the optimization formulation.

	Number of coaches	Operating cost (YEN)	Revenue (YEN)	Profit (YEN)
Baseline	10,10,10,10,10,10,10	1,538,562.50	868,245,230.00	866,706,668.00
Fixed Prices	6,6,6,6,6,6,6	923,137.50	868,245,230.00	867,322,092.50
FixedCoaches-5%	10,10,10,10,10,10,10	1,538,562.50	1,103,365,179.00	1,101,826,616.00
Optimized-5%	6,9,7,7,6,6,6	1,046,650.00	1,103,365,178.50	1,102,318,528.50

Note that in both Baseline and FixedPrices the demand for each trip is identical. Further, the number of seats sold is equal to the demand on the trips on all train lines. Hence, the revenue obtained from ticket sales is identical (see Table 5). However, using the freedom to vary the number of coaches on the train lines enables the optimizer to find a solution where fewer number of coaches are utilised and hence, lower operating costs. As a result, FixedPrices is able to attain a profit improvement of 0.07% over the Baseline. FixedCoaches-5%, which allows price variation while keeping the number of coaches fixed, is able to obtain a 27.2% increase in profit over the Baseline. In FixedCoaches-5% the price for all trips is reduced by 5% in order to increase the demand. The increased number of tickets contributes to the increase in revenues and profits. Finally, Optimized-5% is able to further improve the profit margin by reducing the operating costs. Note that the revenue in the cases of FixedCoaches-5% and Optimized-5% are identical even though fewer coaches are allocated to the train lines in Optimized-5%. The optimal prices identified in FixedCoached-5% and Optimized-5% are identical and hence, the demand that manifests are also identical in both cases. This reinforces the remark made earlier that the number of seats available in Baseline exceeds the demand.

The number of coaches on certain train lines is set to the lower limit (\underline{u}_t) at the optimal solution for Optimized-5% (refer Table 5). This seems to suggest that some train lines may not be necessary. To test this hypothesis, we modified the lower bound on the number of coaches to $\underline{u}_t = 0$ for all $t \in \mathcal{T}$. With this changed bound, we identify an optimal solution where the number of coaches on the lines are 13,0,0,6,0,0,0. In other words, 13 coaches are assigned to Tokyo–Shin-Osaka Nozomi which runs 56 times per day and six coaches are

assigned to Tokyo–Hakata Nozomi which runs 33 times per day. All other train lines are eliminated. Despite the fewer train lines, the revenue is identical to that reported for Optimized-5% in Table 5. However, the operating cost is 527,506 which amounts to a reduction of 50% from the operating costs for Optimized-5% in Table 5. This result provides further evidence that indeed the capacity in Baseline far exceeds the demand.

In the next set of experiments, we study the scalability of our approach when the number of price-demand pairs are increased. As noted earlier, our intention is to work directly with the price-demand data pairs instead of the estimated logit function. The approach of discretizing the prices to obtain the price-demand is advocated in the absence of such raw-data. We simulate the case of increased price-demand pairs by increasing the number of price discretization points (n_{disc}). Table 6 present results from optimizing the number of coaches and prices for different number of discretization of the price $n_{disc} \in \{5,10,20,40\}$. These approaches are respectively called Optimized-5%, Optimized-10%, Optimized-20% and Optimized-40%.

Table 6: Effect of number of price discretization on the solution.

	Number of coaches	Operating cost (YEN)	Revenue (YEN)	Profit (YEN)
Optimized-5%	6,9,7,7,6,6,6	1,046,650	1,103,365,178.50	1,102,318,528.50
Optimized-10%	9,6,6,9,6,6,6	1,078,237.50	1,248,444,333	1,247,366,095.50
Optimized-20%	10,6,6,11,6,6,6	1,150,593.75	1,308,000,398.20	1,306,849,804.45
Optimized-40%	10,6,6,11,6,6,6	1,264,262.50	1,308,123,101.90	1,306,972,508.15

Table 6 shows that the profit improves as additional price discretizations are considered. Optimized-20% identifies an optimal solution with 50.8% improvement in profit over Baseline. The improvement in profit for Optimized-40% is negligible when compared with Optimized-20%. The size of ILPs increases as the n_{disc} is increased. Optimized-5% has 983 variables and 869 constraints while Optimized-40% has 6023 variables and 5909 constraints. Optimized-40% is solved in 0.2 seconds. Hence, solving ILPs with large number of price-demand pairs does not present a computational bottleneck for the proposed approach. Fig. 1 plots the percentage reduction in the prices for the different trips in the solution obtained from Optimized-10%, Optimized-20% and Optimized-40%. The prices are uniformly reduced by 10% for all trips in the solution for Optimized-10%. The solution to Optimized-20% exploits the difference in the price-demand elasticities for the trips to obtain an improved solution. The price for all trips are reduced by at least 13%. The trips Nagoya–Kyoto (#16 on x-axis of Fig. 1), Nagoya–Shin-Osaka (#17 on x-axis of Fig. 1), Nagoya–Shin-Kobe (#18 on x-axis of Fig. 1), Kyoto–Okayama (#24 on x-axis of Fig. 1), Shin-Osaka–Okayama (#28 on x-axis of Fig. 1), Shin-Kobe–Okayama (#31 on x-axis of Fig. 1), Okayama–Hiroshima (#34 on x-axis of Fig. 1) and Okayama–Hakata (#36 on x-axis of Fig. 1) all see a price reduction of 20% in the solution to Optimized-20%. In the solution to Optimized-40% the price reduction is identical to that for Optimized-20% for most trips. The difference occurs for those trips that reduced the price by 20% in the solution to Optimized-20%. These trips are further reduced by 1–2% in the solution to Optimized-40%.

5 CONCLUSION

We propose a data-driven approach for revenue maximization optimization by determining jointly the train sizes and prices for a single fare class over multiple legs and train lines. The resulting optimization problem belongs to the class of ILPs. The ILP is shown to be



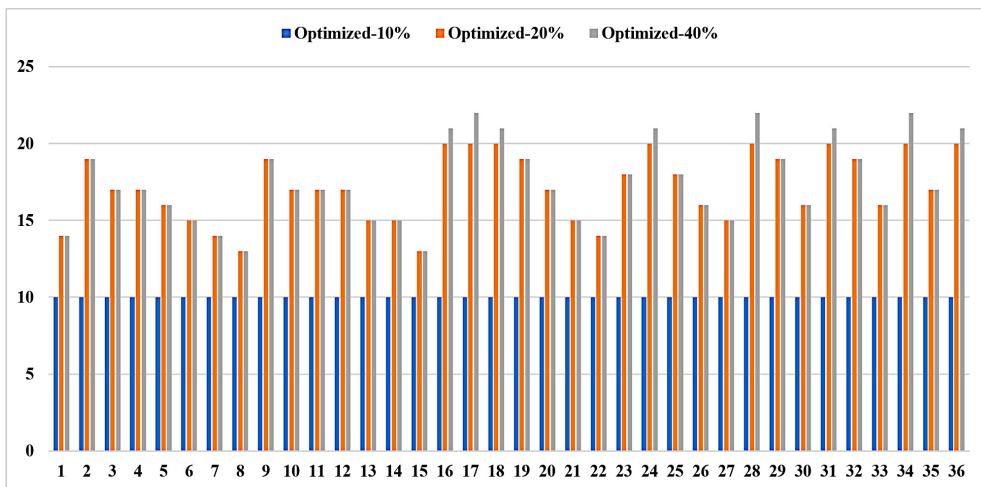


Figure 1: Plot of percentage reduction in price at the optimal solution for the different trips (trips on the x-axis are numbered in the order listed in Table 4).

computationally tractable through a numerical example. The numerical experiments demonstrate that the proposed approach is scalable in the number of price-demand pairs and is effective in improving the revenue by selling increased number of tickets at lower prices. In contrast to existing approaches for joint price and seat allocation that are nonlinear and ignore integral nature of decision variables, our approach shows that joint price and seat allocations can be effectively solved using an ILP formulation.

ACKNOWLEDGEMENT

The second author acknowledges discussions with Professor David Bergman on the connection of discretized logit curves with the data-driven approach.

REFERENCES

- [1] Glover, F., Glover, R., Lorenzo, J. & McMillan, C., The passengers-mix problem in the scheduled airlines. *Interfaces*, **12**, pp. 73–79, 1982.
- [2] Belobaba, P.P., Airline yield management: An overview of seat inventory control. *Transportation Science*, **21**, pp. 63–73, 1987.
- [3] Davis, P., Airline ties profitability to yield management. *SIAM News – Mathematics that Counts*, **27**, 1994.
- [4] Weatherford, L.R. & Bodily, S.E., A taxonomy and research overview of perishable-asset revenue management: Yield management, overbooking and pricing. *Operations Research*, **40**(5), pp. 831–844, 1992.
- [5] McGill, J.G. & van Ryzin, G., Revenue management: Research overview and prospects. *Transportation Science*, **33**(2), pp. 233–256, 1999.
- [6] Talluri, K.T. & van Ryzin, G.J., *The Theory and Practice of Revenue Management*, Kluwer Academic Press: New York, 2004.
- [7] Strasser, S., The effect of yield management on railroads. *Transportation Quarterly*, **50**, pp. 47–55, 1990.

- [8] Ciancimino, A., Inzerillo, G., Lucidi, S. & Palagi, L., A mathematical programming approach for the solution of the railway yield management problem. *Transportation Science*, **33**(2), pp. 168–181, 1999.
- [9] You, P.-Y., An efficient computational approach for railway booking problems. *European Journal of Operations Research*, **185**(2), pp. 811–824, 2008.
- [10] Armstrong, A. & Meissner, J., Railway revenue management: Overview and models. Technical report, Lancaster University Management School, 2010.
- [11] Hetrakul, P. & Cirillo, C., A latent class choice based model system for railway optimal pricing and seat allocation. *Transportation Research Part E*, **61**, pp. 68–83, 2014.
- [12] Wang, X., Wang, H. & Zhang, Z., Stochastic seat allocation models for passenger rail transportation under customer choice. *Transportation Research Part E*, **96**, pp. 95–112, 2016.
- [13] Zhao, X. & Zhao, P., A seat assignment model for high-speed railway ticket booking system with customer preference consideration. *Transportmetrica A: Transport Science*, **15**(2), pp. 776–806, 2019.
- [14] Zhai, X., Zhao, J. & Chen, Q., Optimization of the assignment of tickets for railway networks with large passenger flows. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transport*, **232**(2), pp. 632–642, 2018.
- [15] CPLEX Optimizer, www.ibm.com/analytics/cplex-optimizer. Accessed on: 5 Feb. 2020.
- [16] Gurobi, www.gurobi.com/. Accessed on: 5 Feb. 2020.
- [17] FICO Xpress Solver, www.fico.com/en/products/fico-xpress-optimization. Accessed on: 5 Feb. 2020.
- [18] Ministry of Land, Infrastructure, Transport and Tourism, Survey of mainline passenger flow. www.mlit.go.jp/common/001193645.xlsx. Accessed on: 9 Feb. 2020.
- [19] West Japan Marketing Communications Inc., www.jcomm.co.jp/transit/price/databook/pdf/19D_41.pdf. Accessed on: 9 Feb. 2020.
- [20] JR Outing Net, www.jr-odekake.net/. Accessed on: 15 May 2019.

