

Solving a Dynamic Revenue Management Problem of Airlines by Bid Price Learning and Approximate Policy Iteration

Kuancheng Huang^{1,*}
Pei-Chen Chung²

*Corresponding author

¹ National Yang Ming Chiao Tung University, Hsinchu, Taiwan, Professor, kchuang@cc.nctu.edu.tw

² National Yang Ming Chiao Tung University, Hsinchu, Taiwan, Undergraduate Student, christinechung.mg09@nycu.edu.tw

ABSTRACT

This paper presents a dynamic pricing framework for airline revenue management, leveraging simulation-based Approximate Policy Iteration (API) technique to optimize pricing and capacity control decisions. The framework incorporates a price-sensitive multinomial logit (MNL) function for customer choice modeling to learn the bid prices representing seat opportunity cost to guide pricing and capacity control decisions. We compare two policies: (1) typical capacity control that optimizes product offer sets under fixed prices, and (2) a dynamic pricing policy that jointly optimizes both product offer sets and associated prices. The generation of the dynamic pricing policy uses epsilon-greedy exploration and neighborhood local search for offer set optimization, and constrained linear regression to estimate the value function in the dynamic programming (DP) model. Simulation results show that the dynamic pricing policy significantly increases revenue.

Keywords: Revenue management, dynamic pricing, dynamic programming, bid price learning, approximate policy iteration.

INTRODUCTION

Many past airline Revenue Management (RM) studies have focused on the capacity control associated with the fare classes, the products designed to exercise market segmentation via imposing different prices and specific terms for the same cabin seats. Typical RM control policies are implemented with pre-determined prices, which remain fixed during the selling process. However, customer purchase decisions are often price-sensitive over time, and ignoring this can lead to suboptimal control and significant revenue loss.

With growing computational power and the availability of customer data, dynamic pricing has been applied by many airlines, especially among low cost carriers (LCCs). Williams (2022) empirically showed that dynamic pricing in U.S. airline markets improves both airline profits and overall welfare compared to uniform pricing. On the modeling side, as a foundational work, Maglaras and Meissner (2006) first introduced a dynamic pricing framework using a deterministic fluid model. However, a key limitation of their approach was that the developed model does not incorporate customer choice behavior. In this study, we investigate the value of dynamic pricing with respect to typical capacity control in a single-leg airline RM problem. Specifically, we want to evaluate: **How much more revenue can be gained by introducing dynamic pricing?**

PROBLEM FORMULATION

By relaxing the assumption of demand independence, Gallego *et al.* (2004), Talluri and van Ryzin (2004), and Liu and van Ryzin (2008) developed initial models for capacity control under discrete choice frameworks, allowing for customers' substitution among products with varying degrees of preference. Building upon these models, we consider a single-leg airline RM problem with a fixed seat capacity C_0 and a selling horizon of T discrete time periods. In this context, a product refers to a fare class offered at a specific price with associated restrictions (e.g., refundability, advance purchase rules, and change penalties). In each period, at most one customer arrives, with a probability λ . Upon arrival, the customer decides whether to purchase an available product (fare class) or opt out, based on their utility evaluation. The utility of product j priced at r_j is modeled using a price-sensitive linear function:

$$u_j(r_j) = b_j + \beta \cdot r_j \quad (1)$$

where b_j is the base utility for product j , and β is the price sensitivity coefficient (negative to reflect diminishing utility for higher prices). At each decision period, the operator selects an offer set \mathcal{S} , a subset of products (fare classes) to be made available to the customer. This offer set is the key decision in the RM model, in which the customer's choice is modeled using a Multinomial Logit (MNL) function. The probability that a customer selects product $j \in \mathcal{S}$ is as (2), given that u_0 is the utility of the no-purchase option, and $p_0(\mathcal{S})$ is the no-purchase probability:

$$p_j(\mathbf{S}) = \frac{e^{u_j}}{e^{u_0} + \sum_{i \in \mathbf{S}} e^{u_i}} \quad (2)$$

Let $V_t(c)$ denote the optimal expected revenue-to-go at time t (i.e., from t to T) with the remaining capacity c . Define the marginal value of capacity as $\Delta V_t(c) = V_t(c) - V_t(c-1)$, which can be thought as the opportunity cost of a seat. The objective is to determine the offer set in each period to form the policy that maximizes the total expected revenue represented by $V_1(C_0)$, where C_0 is the capacity at the beginning of the selling horizon. With pre-determined prices, the value function is represented by the Bellman equation in dynamic programming (DP) in (3), with the notations summarized in Table 1.

$$V_t(c) = \max_{\mathbf{S}} \left\{ \sum_{j \in \mathbf{S}} \lambda \cdot p_j(\mathbf{S}) \cdot (r_j + V_{t+1}(c-1)) + (\lambda \cdot p_0(\mathbf{S}) + 1 - \lambda) \cdot V_{t+1}(c) \right\} \quad (3)$$

To extend this framework to a dynamic pricing setting, where both offer sets and prices are decisions, this study further develops the DP formulation in (4), with some additional notations related to pricing decisions.

$$V_t(c) = \max_{\mathbf{S}, \mathbf{r}} \left\{ \sum_{j \in \mathbf{S}} \lambda \cdot p_j(\mathbf{S}, \mathbf{r}) \cdot (r_j + V_{t+1}(c-1)) + (\lambda \cdot p_0(\mathbf{S}, \mathbf{r}) + 1 - \lambda) \cdot V_{t+1}(c) \right\} \quad (4)$$

Table 1: Summary of Notations

Notation	Description
C_0	Initial seat capacity
T	Selling horizon (number of time periods)
λ	Customer arrival probability in each period
j	Index of a product (fare class)
\mathbf{S}	Offer set of products
r_j	Price of product j
\mathbf{r}	Price vector for all products
$u_{j(r_j)}$	Utility of product j at price r_j
b_j	Base utility of product j
β	Price sensitivity coefficient ($\beta < 0$)
u_0	Utility of the no-purchase option
$p_j(\mathbf{S})$	Choice probability of product j under offer set \mathbf{S}
$p_0(\mathbf{S})$	Choice probability of the no-purchase option under offer set \mathbf{S}
$V_t(c)$	Expected revenue from time t with remaining capacity c
π_t	Time-dependent bid price (opportunity cost per unit of capacity)
θ_t	Intercept estimate in the value function approximation
$\bar{V}_t(c)$	Simulated revenue from time t with remaining capacity c

As illustrated by (3) and (4), one of our key research objectives is to evaluate and compare the following two policies, for which decisions are guided by the marginal value (opportunity) of capacity or the time-dependent bid prices derived by the procedure presented in the next section. **Typical Capacity Control** uses fixed prices and, for each time period, focuses solely on optimizing the offer set, the subset of products made available to customers with pre-determined prices. **Dynamic Pricing Policy** dynamically makes the decision over the offer set and the associated prices in each time period.

To quantify the revenue improvement gained from jointly optimizing both product availability and pricing, we evaluate how incorporating price sensitivity can yield better performance in revenue outcomes through simulation-based comparisons.

SOLUTION METHOD

Past RM studies have used various solution methods to tackle different kinds of RM problems. Among them, Gosavi *et al.* (2002) was the first to apply reinforcement learning (RL) techniques to a realistic RM problem based on a semi-Markov decision process, focusing on capacity and overbooking control under fixed pricing. To incorporate dynamic pricing settings, we design the solution approach by extending the simulation-based Approximate Policy Iteration (API) framework proposed by Koh (2017), which aligns with the development trend of RL methodologies for RM problems. The goal is to iteratively learn time-dependent bid prices π_t that approximate the marginal value with respect to the remaining capacity and time. The key steps are detailed below:

Step 1: Initialization

We initialize bid prices π_t uniformly using the 25th percentile of all available prices across all products. This initialization provides a baseline estimate of the marginal value of capacity, which will be refined over iterations. Each product $j \in \mathbf{S}$ is

associated with some discrete price levels.

Step 2: Customer Choice via Price-Sensitive MNL

Customer decisions are modeled using an MNL model with price sensitivity. Given an offer set \mathbf{S} , the utility of product j is defined as in (1), and the no-purchase option has utility u_0 . Based on these utilities, the choice probability for each product is calculated using (2).

Step 3: Epsilon-Greedy Exploration Strategy

When setting the offer set based on the current values of the bid prices π_t , we use an ε -greedy strategy to balance exploration and exploitation, with the exploration rate ε decreasing over iterations. For exploration, as a simple rule, any fare class with a price r_j larger than π_t is offered. On the other hand, for exploitation, a neighborhood search is performed to optimize the offer set.

Step 4: Neighborhood Search

Starting from including all products at their lowest prices, the algorithm iteratively explores small changes, implemented by adding/removing products or adjusting prices, and selects the option with the highest expected revenue. This process repeats until no further improvement is found, ensuring local convergence.

Step 5: Simulating Booking Processes

A large number of iterations simulating customer booking processes (or paths) are performed to collect training data. For each simulated booking path, at each time period, if a customer arrives, an offer set (and the associated pricing decision for the dynamic pricing policy) is selected as explained in the previous two steps. The customer makes a choice based on the probability from the MNL function, and revenue is recorded by selling a seat, except for the no-purchase situation. These simulations generate training revenue data, *i.e.*, $\tilde{V}_t(c)$, to approximate the value function and estimate the bid prices.

Step 6: Bid Price Update via Constrained Regression

Using the training data from the simulation, we estimate the value function according to the linear approximate function in (5). We fit the parameters θ_t and π_t using least squares with the monotonicity and feasibility to ensure that the marginal value of capacity diminishes over time and remains bounded.

$$\tilde{V}_t(c) = \theta_t + \pi_t \cdot c \quad (5)$$

Step 7: Smoothing Across Iterations

To reduce noise, we use exponential smoothing when updating bid prices and model parameters. The smoothing parameters decrease over iterations to ensure stable convergence.

TEST EXAMPLES AND RESULTS

We consider a case of a single-leg flight, which resembles the setting of LCCs, normally without too many transfer passengers. The airline offers four products (fare classes), each with five discrete price levels. The booking horizon consists of $T = 80$ periods, with the arrival probability $\lambda = 0.8$. The product base utilities are set as $u_1 = 7.0$, $u_2 = 6.5$, $u_3 = 6.0$, $u_4 = 5.5$.

To evaluate the performance of the API-based dynamic pricing approach, we perform the sensitivity analyses on the flight capacity C_0 and the price coefficient β in (1). In each setting, the DP model is solved to optimality to generate the best control policy for benchmarking, and the API policy is trained for 30 iterations with 1500 simulations per iteration. Monte Carlo simulations (1000 runs) are used to evaluate the average revenue for each policy. We compare the following four control policies:

DP-CapC: A typical capacity control strategy that optimizes the offer set under fixed prices, serving as a traditional benchmark.

API-DynP: A simulation-based learning policy that jointly optimizes product availability and prices using API.

DP-DynP: The optimal policy from the exact DP model that simultaneously determines offer sets and prices.

FCFS: A baseline, where customers purchase based on seat availability, basically represents the case of no RM control.

The numerical results for cases with different flight capacities, C_0 , are summarized in Table 2, with the price-sensitivity coefficient fixed at $\beta = -0.002$. Setting different capacities leads to varying supply-demand ratios, a crucial factor affecting the potential benefits of applying RM.

The results confirm the effectiveness of the proposed API-DynP policy, whose learned bid prices leading to near-optimal strategies with average revenues falling within 0.1% of the exact DP benchmark (DP-DynP). Furthermore, this dynamic pricing approach demonstrates significant value over traditional method, improving revenue by over 9% compared to the optimal fixed-price control policy (DP-CapC). This highlights the substantial benefit of jointly optimizing both prices and product availability. When compared to a naive First-Come-First-Served (FCFS) baseline, all model-based strategies deliver significant revenue gains, underscoring the fundamental importance of optimization in revenue management. At the same time, the revenue percentage

differences between the RM control policies and the FCFS policy decrease as flight capacity increases, indicating reduced benefits under lower demand-supply ratios.

Table 2: Revenue comparison among different control strategies for various capacity levels

C_o	API-DynP	DP-DynP	DP-CapC	FCFS	%(API-DynP - DP-DynP)	%(DP-CapC - DP-DynP)	%(FCFS - DP-DynP)
50	38971.78	38983.34	34991.39	23632.70	-0.03%	-10.24%	-39.38%
60	45050.97	45093.77	40763.05	28231.30	-0.09%	-9.60%	-37.39%
70	45859.56	45916.91	41701.23	30201.90	-0.12%	-9.18%	-34.22%

To further evaluate the robustness and impact of customer price sensitivity, we conducted another sensitivity analysis by increasing the price-sensitivity coefficient β , while keeping the capacity fixed at $C_o = 60$. The results are presented in Table 3.

Table 3: Revenue comparison under different control strategies and price sensitivity coefficient β .

β	API-DynP	DP-DynP	DP-CapC	FCFS	%(API-DynP - DP-DynP)	%(DP-CapC - DP-DynP)	%(FCFS - DP-DynP)
-0.002	45050.97	45093.77	40763.05	28231.30	-0.09%	-9.60%	-37.39%
-0.004	40165.76	40188.65	36580.55	27232.90	-0.06%	-8.98%	-32.24%
-0.006	34291.52	34306.01	31319.13	26452.00	-0.04%	-8.71%	-22.89%

For this part of the analysis, the control policy based on the learned bid prices (API-DynP) again achieves revenue very close (0.1%) to that of the optimal control policy (DP-DynP), validating the effectiveness of the developed method. Regarding the performance advantage of the dynamic pricing approach over the traditional fixed-price control policy (DP-CapC), the revenue differences remain significant, still around 9%. However, it is interesting to note that the advantage level decreases (from 9.60% to 8.71%) when the parameter β becomes more sensitive (from -0.002 to -0.006). This seemingly counterintuitive result challenges the simple notion that the more price-sensitive the passengers, the more beneficial the dynamic pricing approach should be.

However, the correct interpretation should consider the overall potential of applying RM by examining the percentage gaps between the dynamic pricing policy and the FCFS policy, i.e., the last column of Table 3. When passengers become more price-sensitive, it becomes more difficult for airlines to apply RM techniques to sell high-fare tickets, substantially limiting the room for RM. As shown in Table 3, the RM potential decreases from 37.39% to 22.89%, while the revenue gap between the dynamic pricing method and the traditional fixed-price control policy only slightly decreases, by about 1%. The percentage of captured revenue highlights the necessity of adopting dynamic pricing control in markets with high price sensitivity.

CONCLUSIONS

This paper addresses an airline revenue management problem by designing a dynamic modeling framework that incorporates a price-sensitive MNL model to better capture customer behavior. The main contributions include the extension of the value function in existing dynamic programming models to account for pricing decisions, and the development of a solution approach that jointly optimizes product availability and dynamic pricing. Numerical results demonstrate the promising revenue benefits of dynamic pricing across varying demand and capacity conditions, with clear performance gains over traditional fixed-price policies. Future research could explore more accurate customer choice models, integration of real-time demand learning, and applications to more complex airline network structures and competitive environments.

REFERENCES

- Gallego, G., Iyengar, G., Phillips, R., & Dubey, A. (2004). Managing flexible products on a network. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3567371>
- Gosavi, A., Bandla, N., & Das, T. K. (2002). A reinforcement learning approach to a single leg airline revenue management problem with multiple fare classes and overbooking. *IIE Transactions*, 34(9), 729-742. <https://doi.org/10.1023/A:1015583703449>
- Koch, S. (2017). Least squares approximate policy iteration for learning bid prices in choice-based revenue management. *Computers & Operations Research*, 77, 240-253. <https://doi.org/10.1016/j.cor.2016.07.015>
- Liu, Q., & van Ryzin, G. (2008). On the choice-based linear programming model for network revenue management. *Manufacturing & Service Operations Management*, 10(2), 288-310. <https://doi.org/10.1287/msom.1070.0169>
- Maglaras, C., & Meissner, J. (2006). Dynamic pricing strategies for multiproduct revenue management problems. *Manufacturing & Service Operations Management*, 8(2), 136-148. <https://doi.org/10.1287/msom.1060.0105>
- Talluri, K., & van Ryzin, G. (2004). Revenue management under a general discrete choice model of consumer behavior. *Management Science*, 50(1), 15-33. <https://doi.org/10.1287/mnsc.1030.0147>
- Williams, K. R. (2022). The welfare effects of dynamic pricing: Evidence from airline markets. *Econometrica*, 90(2), 831-858. <https://doi.org/10.3982/ECTA16180>