Bayesian Reinforcement Learning to Optimize Paid Ancillary Revenue in the Airline Industry

Kevin Duijndam* Ger Koole† Rob van der Mei‡

January 24, 2025

Abstract

To optimize the pricing of paid ancillary seats, we adopt a revenue management approach that optimizes over the capacity of these seats while accounting for unknown underlying model parameters. We test various models against a simulation model to assess the performance against wide-ranging input parameters.

We demonstrate that using a Bayesian exponential demand model to describe the relationship between price and seats sold, combined with a Bayesian reinforcement learning approach to estimate its parameters, outperforms other approaches. By using a relatively simple demand model with a limited number of parameters, updating in a Bayesian manner, and in one step estimating demand parameters to directly use for price optimization, the model is quickly able to perform well across a wide range of demand scenarios.

1 Introduction

Recent advances in how the offer and order systems in airlines work have made it possible to dynamically price ancillary paid seats, and there has been growing interest in optimizing the pricing of those seats (Mumbower, Hotle, and Garrow 2022). When talking about dynamic pricing of paid seats, we specifically mean the price determination of ancillary seats onboard the aircraft. These seats have distinct characteristics: they have limited supply compared to the total number of passengers on board the aircraft, and are of no value anymore after the aircraft departs.

*Department of Mathematics, Section Stochastics, Vrije Universiteit, Amsterdam, The Netherlands and KLM Royal Dutch Airlines

†Department of Mathematics, Section Stochastics, Vrije Universiteit, Amsterdam, The Netherlands

‡Centre for Mathematics and Computer Science (CWI), Department of Probability and Stochastic Networks, Amsterdam, The Netherlands and Department of Mathematics, Section Stochastics, Vrije Universiteit, Amsterdam, The Netherlands

Airline ancillary revenues have grown in importance over recent years, growing from 6.0% of global airline revenues in 2013 to 15.0% of global airline revenues in 2022 (see OAG 2023 for more detail). The airlines with the highest percentage of ancillary revenue as a percentage of total are usually the low cost carriers, with up to 56% of revenue coming from ancillary products (in 2021, by Wizz Air). But full service carriers have been adjusting their models as well, unbundling their products and trying to tailor to more specific customers' preferences, and with that, willingness-to-pay. Boosting the ancillary revenues is seen as a key lever to increase profitability. See for instance Babić, Ban, and Bajić 2019 for more details.

We have developed and tested various approaches to optimize paid seat pricing, building up to a Bayesian exponential model to describe the relation between price and seat sales. Our goal here is to optimize over the capacity of available seats in the aircraft. This is different from most dynamic pricing approaches, which try to find the optimal price point based on customer characteristics. Instead of using a customer choice model, we take an approach that is closer to revenue management. There is usually little data available to estimate the relationship between price and demand for these types of seats. So, in this work, we test these models against various demand assumptions using a simulation model, and show the performance of these models.

Objective and Contributions. The primary goal of this paper is to estimate demand for paid ancillary seats under limited data and significant uncertainty. We identify gaps in the literature where existing methods assume rich historical data or do not account for inventory constraints of specific seat categories. Our contributions are:

- We propose a *Bayesian exponential demand model* for seat sales, tailored to low-data settings.
- We propose a model that optimizes over the capacity of available seats, instead of per customer, utilizing the full capacity of the aircraft.
- We compare multiple methods (Q-Learning, linear regression variants, binomial model) within a unified simulation framework, showing how each handles exploration-exploitation differently.
- We show that with stylized assumptions, our Bayesian approach quickly learns seat-pricing strategies that capture a high percentage of theoretical maximum revenue.

2 Related Work

There is a strong link to both more traditional revenue management used for ticket pricing, and dynamic pricing in general in this specific area. In this paper, we follow the definition as proposed in Wittman and Belobaba 2019, in which dynamic pricing is defined as 'Firms practice *dynamic pricing* when they charge

different customers different prices for the same product, as a function of an observable state of nature'.

2.1 Ticket revenue management

A general framework for ticket revenue management usually has two steps. A first step is to model willingness-to-pay (WTP) curves and arrival rates of types of customers and to estimate their parameters (based on historical data). The second step is then to optimize the network based on the estimated model. See for example Belobaba 1989 for a starting point in this. In Gosavi 2004, this is done more directly with a model-free reinforcement learning approach by solving an average reward Markov (and semi-Markov) decision process, based on Q-Learning that approximates the value function. In Otero and Akhavan-Tabatabaei 2015 a different approach is presented, where a stochastic dynamic pricing model is used to optimize ticket pricing. A phase-type distribution is used to model inter-arrival times of customers, and the probability that a customer will buy a ticket. These inter-arrival times are modeled explicitly, in order to calculate the results.

2.2 Airline ancillary pricing

There is less research done on airline ancillary pricing than on ticket pricing, however also in this domain there are quite a few papers of interest. Several papers estimate customer willingness-to-pay through surveys, using individual customer characteristics to predict the best price for that individual customer. See for instance Warnock-Smith, O'Connell, and Maleki 2017, Shukla et al. 2019, and Zhao, Cui, and Cheng 2021. In all these studies, surveys are conducted to learn about the relation between customer characteristics (such as route, departure / return dates, number of passengers in the booking, etc.) and willingness-to-pay.

In Ren, Pan, and Jiang 2022 this is done without surveys, but by estimating the parameters by dividing the customers in three distinctive groups using a latent class conditional logit model. See Kummara et al. 2021 for a more practical implementation approach, using gradient boosting to estimate ancillary prices. In this paper, inventory limits are not considered.

In general these approaches are more suited to products without a hard inventory limit (like checked baggage), than it is for paid seats, as it is not taken into account how many of the seats will be sold. Therefore, the model does not necessarily optimize the revenue that can be achieved on a flight.

Both Ødegaard and Wilson 2016; Shao and Kauermann 2020 optimize ticket and ancillary pricing together in one optimization problem, mainly focused on bundles. This is done through dynamic programming and statistical regression methods. Besides, in Wang, Wittman, and Bockelie 2021, extensions are made on an earlier customer choice model that optimizes both ticket and ancillary pricing together, also from the perspective of assortment optimization (see also Wittman and Belobaba 2019). This is also done in Wilson and Ahmed

2024, where the problem of setting a price for a combination of a ticket and multiple ancillary products is considered. Interesting from that paper is the time-dependent arrival notion where certain passenger types arrive at certain times.

2.3 Dynamic pricing

Vast literature exists on dynamic pricing in general. In this paper, we focus on a reinforcement learning approach, so we narrow the selection down to papers that also take such an approach. More often, a form of Q-Learning, so a model-free approach is used, see for instance Dogan and Güner 2015; Mishra, Mosutafa, and Ito 2020. But the focus is on infinite horizon, multi-retailer setting or an auction type environment, which is different from a fixed inventory, single seller, type setting. The Q-matrix can also be approximated in these settings with a neural network to learn faster.

An interesting comparison is with Bastani, Simchi-Levi, and Zhu 2022, where a Bayesian-approach is also used, with a so-called meta-prior that learns across experiments, a methodology that is also used here. There, the main focus is on transfer-learning across pricing experiments, but less on which model to use to optimally determine a price. Another interesting comparison is with Yang, Chu, and Wu 2022 in which a finite inventory pricing problem is modeled. An LSTM-network is used to estimate the WTP-distribution, and then an MDP pricing model is used to generate an optimal pricing strategy. There is also an interesting comparison with Harrison, Keskin, and Zeevi 2012, who also uses a Bayesian approach, where a seller offers prices sequentially to customers, which is the same setting as we have. In that paper, a binary prior model is used, in which upfront one of two possible demand models is applicable (unknown to the seller which one). Bayesian reinforcement learning is used to converge to a belief as to which of the possible demand models is active. As mentioned in that paper also, this is a more simplistic view, where usually in practice there are not two possible types of demand. Luo, Sun, and Liu 2024 also provides an interesting insights, where a linear bandit framework is used with an upper-confidence bound approach to balance exploration and exploitation, with unknown noise parameters. In this paper a customer choice model is also considered with a linear valuation function.

A very different approach is used in Boer and Zwart 2014, called Controlled Variance Pricing, to optimally balance learning and gaining revenue. No model is assumed between price and willingness-to-pay. Instead, the approach guarantees a minimum (shrinking over time) variance of the prices tested. With that, the policy ensures that enough data is gathered to test a wide enough range of prices, while converging to the optimal price.

For a complete overview of various approaches in dynamic pricing in a wider range of settings (e.g., also without inventory constraints, or different approaches), see Den Boer 2015.

2.4 Our contribution

Instead of using a completely model-free approach, we use relatively simple models with a few parameters to facilitate faster learning. This approach leaves various underlying mechanisms like inter-arrival times implicit, does not require accounting for O&D complexities, and combines the estimation and optimization steps often seen in ticket revenue management approaches in one. This gives an improvement over the current literature by being practically more usable.

While estimating willingness-to-pay through surveys can work well for products without hard inventory limits (like check-in baggage), our approach does take into account the inventory of paid seats. This is important to optimize the total revenue from the available number of seats. Our approach is closer to a revenue management approach, trying to make optimal use of the finite inventory, instead of modeling customer characteristics. This is to our knowledge not done yet in current literature and valuable in practice.

Compared to some of the dynamic pricing literature referenced in previous sections, we take an approach that needs less known demand parameters, while at the same time we use a simpler model that facilitates faster learning, as there are fewer parameters to learn (instead of the approach in Yang, Chu, and Wu 2022. But we do consider important problem characteristics like inventory constraints, to optimally set a price for the full paid ancillary seat inventory (instead of setting an optimal price from a customer choice optimization perspective like in for instance Luo, Sun, and Liu 2024 or Wang, Wittman, and Bockelie 2021). Besides, we use a similar mechanism as in Bastani, Simchi-Levi, and Zhu 2022 to enable transfer-learning between flights, but together with an optimized pricing function which has less focus in that paper. This is to our knowledge also not done in other literature currently, and makes the model practically more usable - although comes at the cost of not allowing to prove any optimality bounds.

3 Methodology

There are two main parts to consider: the general demand model, and how this is used to optimize pricing. We will start with more simple models, building up to our final approach and main contribution, a Bayesian exponential model.

3.1 Model Assumptions and Limitations

As far as we know, there is no public data available that describes well the relation between price and demand for these types of seats, so it is not possible to estimate specific distributions that model this relation, with the certainty that that is similar to reality. In this paper, we assume that the number of seats sold at departure of the aircraft is a function of the price of these seats, with a maximum the number of seats that are the actual capacity. We assume there are more than enough potential customers in the aircraft, so that it is not necessary to explicitly model potentially arriving customers. In general in an airline context, this is a safe assumption for specific paid seats, where the aircraft

configuration is generally always such that paid seats $\ll$ regular customers, and it can be assumed that the overall revenue management system will ensure enough customers are on the flight.

We also assume that different flights are similar in patterns, so that learning from one flight can be used on the next. Note that this means we assume a meta-prior distribution that governs the prior distribution for all flights. It does not mean that the distribution for every flight needs to be the same. This transfer learning is explained in more detail in the model in section 3.8. Further work could take into account how this meta-prior can be updated for e.g. seasonality. A practical approach right now could be to have various models for various times of the year if there are relevant differences. It could also be that other characteristics of flights (different aircraft or time of day) are so different that it is not optimal to use the same model across all flights. Since there is no public dataset with real-life paid seat sales available to our knowledge, we could not test these assumptions in practice.

Across all models, we assume for marketing reasons, it is necessary to keep the price fixed for the 'lifetime' of the flight. This means that the model can only choose an initial price, and learns the results of the chosen price after the flight has departed and we know the total number of seats sold. Besides, we assume that for every specific seat in the aircraft, the same price will be charged. Through manual testing on various airline websites, we have found that this is current practice by many large airlines at the time of writing. Further work could investigate both the value of updating the prices throughout the sales process, and differentiating prices within a type of seats, based on where the physical seat is located in the aircraft. Since again, there is no public dataset available to test these assumptions, we can't be certain that optimizing over the inventory is a better strategy than optimizing based on customer characteristics.

3.2 Overview of Approaches

This section introduces several approaches, some of which (e.g., Q-Learning, linear regression, binomial model) serve as benchmarks or baseline methods, and one main new approach (the Bayesian exponential model). We do not intend this as a comprehensive survey; instead, we demonstrate how well-known methods compare to our proposed *Bayesian exponential approach*, highlighting advantages in fast learning and robust performance under uncertainty.

3.3 Q-Learning

The problem can be viewed as a standard Q-Learning problem. In such a setting, the actions are viewed as the price to be used for this flight, and the reward is then the revenue in the end achieved at departure of the flight, which together constitute the Q-table.

The advantage of this demand model is that no relation between price and sales is assumed in advance, so the model is able to learn any form. However, a set of possible prices will have to be defined upfront, and no continuous range

of prices is possible. It is necessary to set both the range, and the granularity of prices, which are important tuning parameters. The optimal price needs to be in the range of possible prices, and the granularity needs to be balanced between quickly finding the right price – which is easier with less granular possible prices, and being close enough to the optimal price – which is easier with more granular possible prices. Another disadvantage is that the model doesn't learn across price points. Every separate price point needs to be tested by the model to learn the resulting seat sales, and no information is transferred between the price points.

3.3.1 Demand model

The demand model is then completely model-free. A Q-table per route is created, with only the possible actions as the price to use for the flight, and as result the revenue obtained in that case. A wide range of prices is used as input. Updating the Q-table is done through the regular update rule

$$Q^{new}(a_t) \leftarrow (1-\alpha) \cdot Q(a_t) + \alpha \cdot \left(r_t + \gamma \cdot \max_a Q(a) \right),$$

with s_t the number of seats available at time t, t is a flight that departs, so every new flight increments t by 1, a_t the price chosen, α the learning rate (here set at 0.995), r_t the revenue obtained, and γ the discount factor (here set at 0.9).

3.3.2 Price determination

In general, the price is then determined by selecting the action with the highest Q-value, which is the exploitation strategy. To have enough exploration as well, the trade-off between exploration and exploitation is in this setting determined through a regular ϵ-greedy approach (here set at 0.01). See section A.1 for more detail on the sensitivity of these parameters.

3.4 Linear regression

3.4.1 Demand model

In this setting, a linear relation is assumed between the price of the seats, and the resulting number of seats sold. So the relation used is

$$s = a - bx \tag{1}$$

with x the price that is chosen, s the seats sold, and parameters a, b that have to be learned. After at least 2 different prices were chosen (at random), a linear fit is done, minimizing the error between the actual seats sold and what is predicted by the model.

An advantage of this approach is that it allows continuous price points, without the need to specify possible prices upfront. Next to that, learning happens faster, because of the assumed linear relation. The model also learns about untested price points, by estimating the parameters a, b.

3.4.2 Price determination

Learning in this case is done by using a regular ϵ-greedy approach to determine the trade-off between exploration and exploitation (set at 0.05, in which case a random price between a predetermined minimum and maximum is chosen).

In case exploitation is chosen, the price that optimizes revenue is chosen by finding the maximum over

$$\text{revenue} = \min(C, s) \cdot x$$

with the expected seats sold s coming from the linear relation, and C the seat capacity of the aircraft. Since prices can be set up to the cent level, this optimization can easily be done numerically. See section A.2 for more detail on the sensitivity of these parameters.

3.5 Constrained linear regression

This setting is very similar to the previous linear regression. However, in this case, we set the constraint that a higher price cannot lead to higher demand.

3.5.1 Demand model

The same linear relation between the price of the seats and the resulting number of seats sold is assumed.

$$s = a - bx$$

However, now with the constraint that $b \geq 0$, which results in a curve that is non-increasing if x increases. After at least 2 different prices were chosen (at random), a linear fit is done, minimizing the error between the actual seats sold and what is predicted by the model, under the condition that $b \geq 0$, using coordinate descent. Due to randomness across flights, it is well possible that choosing a higher price for another flight, still leads to more seat sales, simply because the willingness-to-pay on the other flight is higher. In case that happens, a, b are still set at the point that minimizes the error between actual seats sold and what is predicted by the model, just with the condition that $b \geq 0$. This means that there might still be other choices for a, b that lead to a smaller error.

3.5.2 Price determination

Like in the case of regular linear regression, learning is done through a regular ϵ-greedy approach to determine the trade-off between exploration and exploitation (set at 0.01).

Exactly like in the case with the linear model, in case exploitation is chosen, the price is chosen that optimizes revenue by finding the maximum over

$$\text{revenue} = \min(C, s) \cdot x$$

with the expected seats sold $s$coming from the linear relation. In this case, finding the maximum is also done numerically.

3.6 Bayesian linear regression

3.6.1 Demand model

The same linear relation between seats sold and price can also be used in a Bayesian approach. So still with the model between price set and seat sales as:

$$s = a - bx$$

However, instead of learning a specific value of the parameters a and b, since they are unknown, we assume a distribution for them. Like in the previous approach, we would like to have a curve that is non-increasing if x increases. In order to achieve this, we use a log-normal distribution for b, and a normal distribution for a. Since the log-normal distribution is only positive, this means that the curve can only be non-increasing. We achieve this by setting a normal-inverse-gamma distribution as prior for b. This is the conjugate prior distribution for a posterior with normal distribution with unknown μ and σ, meaning that the posterior distribution for b would be normally distributed and can be analytically computed. In order to set $b \geq 0$, we apply a logarithm to the data to compute the posterior parameters, leading to a log-normal distribution for b. For more detail see for instance Fink 1997.

An advantage of this approach is that we can be smarter about balancing exploration and exploitation. Instead of setting an explicit ϵ-parameter to determine how much exploration is done, we can use the distribution of the parameters. Thompson sampling is used to have a specific value for the a and b parameters (originally coming from Thompson 1933). This means that every time we need to compute the demand distribution, we take a random sample out of the posterior distributions for a and b, and optimize based on those values. By taking a random sample according to the distribution of these parameters, we still test various prices, but we are more likely to test in areas where the probability is 'high' that we test something logical.

3.6.2 Price determination

To have a specific estimate for the seats sold given a chosen price, the expected value of the parameter distributions is used to represent the expected number of seats sold. The price is then optimized by choosing the price that has the highest expected revenue, in the same manner as the two previous linear regression approaches.

3.7 Binomial model

3.7.1 Demand model

All previous methods were regression-based approaches, and here we use different approach. The main idea is that for a specific flight type, every price point has its own probability that a seat will be sold. This can then be interpreted as a binomial experiment, in which there are C independent Bernoulli trials, for

the number of seats available, with a probability p that a seat is sold. So the probability that k seats are sold at departure, is then

$$P(s = k) = \binom{C}{k} p_x^k (1 - p_x)^{C-k},$$

with p_x the probability for price level x, with s the number of seats sold, and C the seat capacity. In this case, we need to define a set of allowed prices $X = \{x_1, x_2, ...\}$. The expectation for the total number of seats sold is then the probability that a seat is sold multiplied by the number of seats.

The main choice is then how to learn p_x, the probability per price point, as input for the binomial experiment. This can be done in a frequentist approach, by keeping track of how often seats were offered at a price point, and how often they were sold. Then the new probability for this price point is simply the division of number of sales by number of offers:

$$p_x = \frac{\#\text{seats sold at price } x}{\#\text{price } x \text{ offered}} \tag{2}$$

A Bayesian approach can also be used with this demand model, which has shown to result in better performance. In that case, a Beta(1,1)-prior is assumed for the p_x parameter, to start with a wide range of possibilities for the probability. This Beta distribution is the conjugate prior distribution for a binomial posterior, hence fits well with this binomial model. Updating the two parameter values of the prior can then simply be done analytically afterwards, by incrementing the number of offers and number of sales for the two parameters of the Beta distribution (see again Fink 1997 for more detail on conjugate pairs).

3.7.2 Price determination

Then the total expected revenue gained by choosing a seat price is

$$\mathbb{E}(\text{revenue}) = \sum_X p_x(\text{seat sold}) \cdot \text{seat price} \tag{3}$$

The advantage of using the Bayesian approach for learning the probabilities is that Thompson sampling can be used to balance the trade-off between exploration and exploitation. Instead of using equation (2) which gives a specific point estimate for the probability of purchase at a price point, we use the beta-distribution. By taking a random sample from this distribution, we explore the range of possible probabilities for this price-point, but are more likely to do so at 'logical' places. Then with these samples, we choose the price point that optimizes the expected revenue according to equation (3). Since in this case we have a set of allowed prices, we simply numerically determine the maximum. After we get the results for this price point, we can again update our belief in the distribution for the probabilities.

3.8 Bayesian Exponential Model

Bayesian updating for demand estimation has been explored in ticket-pricing contexts (Harrison, Keskin, and Zeevi 2012; Bastani, Simchi-Levi, and Zhu 2022). Our exponential model is relatively simpler yet highly flexible, incorporating capacity effects through the logistic-like form. Those elements separately were presented in other contexts before, but this combination, along with the hierarchical prior (meta-distribution), is not found in ancillary seat pricing literature, marking our main methodological contribution.

3.8.1 Demand model

In this case, it is assumed that the demand follows an exponential curve describing the relation between price and demand, like $D = e^{b+ax}$, with D the overall demand, a, b parameters that describe this relation, and x the price chosen. Note that this is also a simplification assuming continuous demand, while in practice demand is discrete. However, since the number of paid seats is significantly lower than the total number of customers on the aircraft, sales of paid seats cannot be higher than the capacity. So in the end we observe sales instead of demand, described by

$$S = \min(C, e^{b+ax}), \tag{4}$$

with S a random variable that describes the sold seats, C the paid seat capacity, and all other parameters as before. See figure 1 for an example with both the potential total demand, and the total sales that would have occurred.

3.8.2 Sales model approximation

Equation (4) is inconvenient in further computations due to the presence of the minimum function. In order to have a simpler function that can be used more easily in further computations, we approximate the sales by

$$S = \frac{C}{1 + e^{-(b+ax)}}, \tag{5}$$

with S sold seats, C seat capacity, x price chosen. a, b are estimated parameters used to fit the model, which are specific to the flight (= route) of interest. With a, b it is possible to model a wide range of options. The curve can be close to linear, with a positive or negative slope, ranging between the maximum capacity and 0. This approximation via a logistic-like function (Equation 5) is a common approach in practice to smooth out the discontinuity arising from $\min(C, e^{b+ax})$. While (4) is piecewise differentiable, we find that the form in (5) eases Bayesian updating and yields robust results in our experiments. In typical seat-capacity ranges (e.g., 5–20 seats), the approximation is quite close.

Figure 2 shows for the same example parameters how the capped demand model to describe actual sales can be approximated with equation (5).

In (5), a, b are unknown parameters which we have to learn, and which might also differ per specific flight. Therefore, we approach these as distributions,

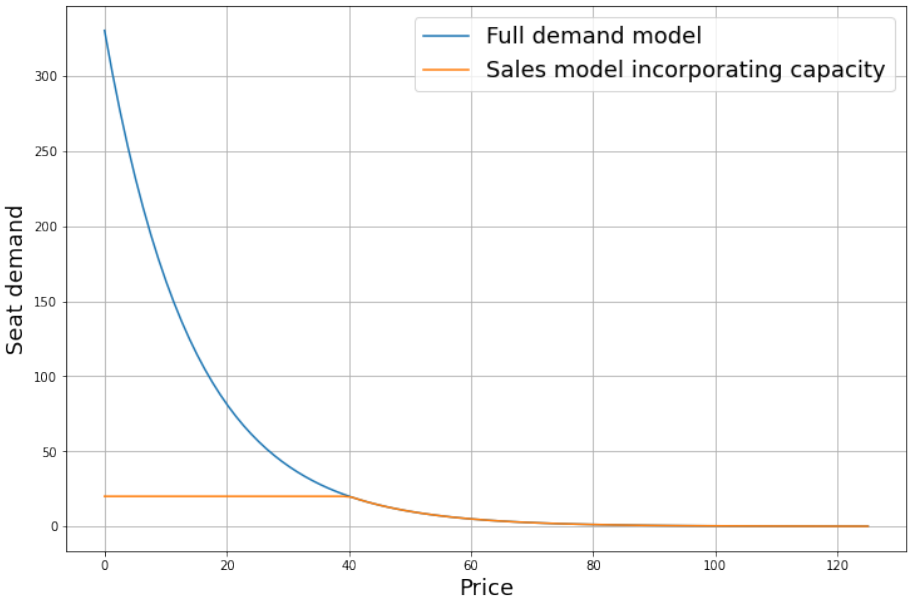


Figure 1: An example of the full demand model with parameters $a = -0.07, b = 5.8$ and $C = 20$.

leading to a Bayesian approach for modeling the relation between price and seats sold. Based on the results of the number of seats sold, a, b are updated in a Bayesian manner. We assume that $a \sim \mathcal{N}(\mu_a, \sigma_a), b \sim \mathcal{N}(\mu_b, \sigma_b)$.

After a flight has departed, we know the actual number of seats that are sold, denoted as s. That means then that

$$-log(\frac{C}{s} - 1) \sim \mathcal{N}(\mu_b + \mu_a \cdot x, \sigma_b^2 + \sigma_a^2 \cdot x) \tag{6}$$

In Equation 6, the ratio $\frac{C}{s} - 1$ is stochastic, which arises from s since this has an underlying distribution of customer willingness-to-pay and how many seats are sold. We then take a log transform to relate the observed s to $b + a\,x$ in a convenient way. This is not a straightforward conjugate prior, therefore, the Metropolis-Hastings (MH) algorithm is used to update the parameters (originally proposed in Hastings 1970). We describe how that is applied in this case, but for a full and detailed background, see for instance Chib and Greenberg 1995. In short, we want to be able to sample from equation (6) in order to estimate the updated $\mu_a, \mu_b, \sigma_a, \sigma_b$, and use that for the following flight. In order to do that, following the MH-algorithm:

1. Set some initial random guess for x^0.
2. Repeat for $j = 1, 2, \ldots, N$.

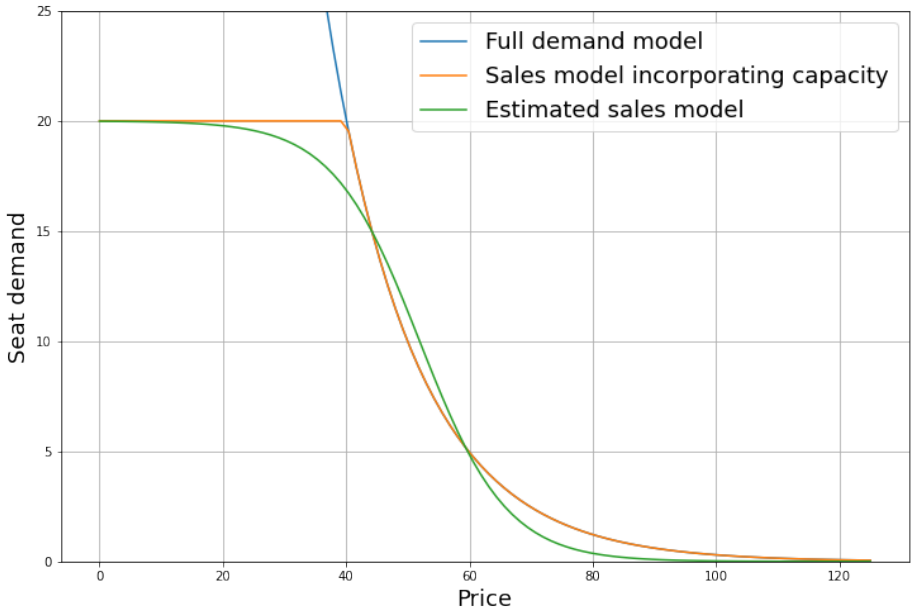


Figure 2: The same demand and sales model, with an approximation by the exponential curve.

3. Generate y from $q(x^j, \cdot) = x^{j-1} + \mathcal{N}(\mu_{MH}, \sigma_{MH})$ and u from $\mathcal{U}(0,1)$.
4. If $u \leq \alpha(x^j, y)$
   - set $x^{j+1} = y$
5. Else
   - set $x^{j+1} = x^j$
6. Return the values $\{x^1, x^2, \ldots, x^N\}$.

where $\alpha(x, y)$ is defined as the minimum of the ratio of the likelihood of the data under parameters x and y, and 1: $\alpha(x, y) = \min(\frac{\mathcal{L}_{data}(x)}{\mathcal{L}_{data}(y)}, 1)$. Since this likelihood is described by equation (6), this can be easily computed. In the practical experiments, we have used $N = 5000$, so 5000 samples to estimate the posterior distribution. We set $N = 5000$ samples in the Metropolis-Hastings algorithm to ensure good mixing and convergence in our posterior estimates. Although fewer samples (e.g., 1000–2000) sometimes suffice, we found 5000 strikes a balance between computational overhead and accuracy in our trials. On a regular laptop (Macbook Pro M1), the algorithm then runs in seconds per flight. If runtime is more constrained, advanced MCMC techniques (e.g., Hamiltonian Monte Carlo, for instance No-U-Turn Sampler, see Hoffman, Gelman, et al. 2014 for

more detail) or fewer iterations can be considered. However, in our practical experiments we have found that the calculation time on a regular laptop was acceptable with this configuration.

We can then assume that across flights, there is a shared structure, with curves that are 'similar' across different routes. This shared learning can be modeled in a hierarchical Bayesian set-up, again assuming that $\mu_a, \mu_b, \sigma_a, \sigma_b \sim$ are normally distributed.

In figure 3 is an example of this relation in practice, based on a few samples of data. The green area shows the 95%-width of the posterior distribution that describes the relation between price and seats sold, which of course depends on the spread of sampled data and the amount of data available.

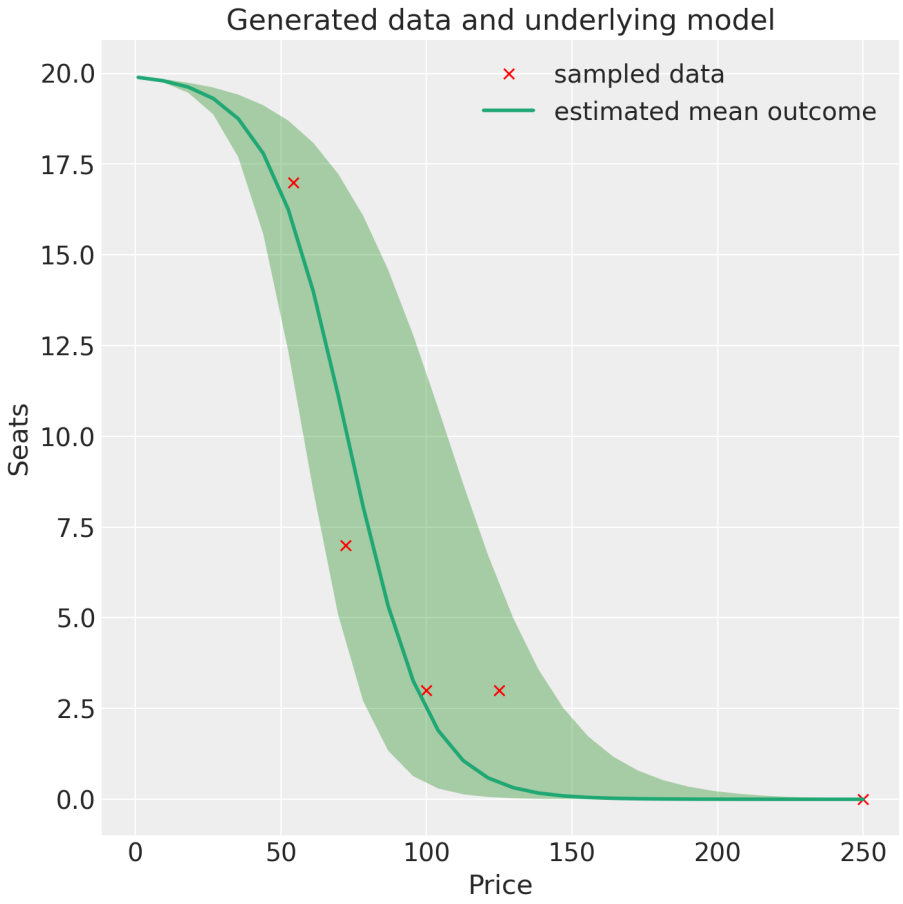


Figure 3: The distribution and mean of the posterior distribution after a few samples from a simulation with a 95%-confidence interval.

3.8.3 Price setting

Based on expert input, or just a wide-ranging distribution if this is not available, a prior distribution can be chosen for the a, b parameters. Using the simulation model from section 4.1, we have found that a normal distribution works well for both these parameters. The mean and variance of these normal distributions are then again from a meta-distribution that is shared across types of flights (as proposed in Bastani, Simchi-Levi, and Zhu 2022). We have found that for this meta-distribution, normally distributed parameters also perform well. So, that leads to $\mu_a \sim \mathcal{N}(\mu_1, \sigma_1), \sigma_a \sim \mathcal{N}(\mu_2, \sigma_2), \mu_b \sim \mathcal{N}(\mu_3, \sigma_3), \sigma_b \sim \mathcal{N}(\mu_4, \sigma_4)$. Figure 4 shows the relation between these distributions. We select normal priors for simplicity and because normal distributions form a flexible family in practice. Although not guaranteed to perfectly match real seat sales, we find that normal priors on $(\mu_a, \sigma_a, \mu_b, \sigma_b)$ offer a good empirical fit over multiple flights. Other priors (e.g., Gamma distributions) could also be used, depending on domain knowledge.

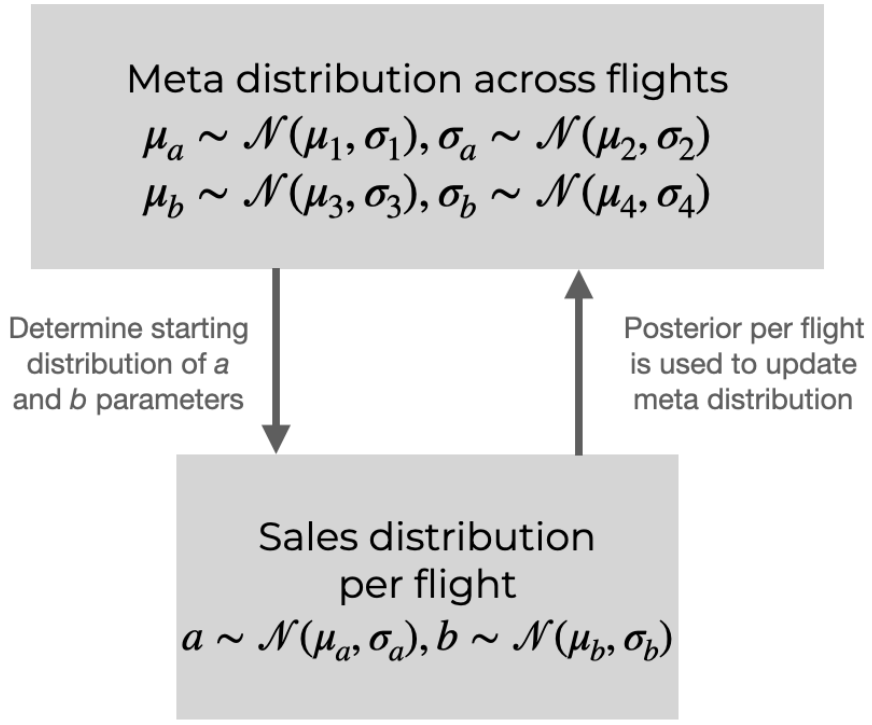


Figure 4: The relation between the meta-distribution across flights, and specific distribution per flight.

Across all flights, based on expert input, a prior is set for $\mu_a, \sigma_a, \mu_b, \sigma_b$. Then, for every flight, Thompson sampling is used to have a specific value for the $\mu_a, \sigma_a, \mu_b, \sigma_b$ parameters (originally coming from Thompson 1933). After that, for that specific flight, using Thompson-sampling again over the a, b parameter distributions, we get a specific value of a, b, and with that we have a direct relation between the price and the seats that are sold at departure of the aircraft. It's then straightforward to choose the price that optimizes expected revenue given that all parameters are known, by finding the maximum over

$$\text{expected revenue} = \text{expected seats sold} \cdot \text{seat price}$$

After departure of the flight, we learn the actual amount of seats sold, and the posterior distribution describing the relation between seats sold and price can be updated. The algorithm here does not have any knowledge about the actual distribution of the demand, but works with the price and resulting seats sold. For the 'next' flight within this type of flights, this updated posterior is then again used as prior.

After all flights have departed, the results for a, b can be used to update the meta-prior that is shared across flight types. Since these are regular normal distributions, it is possible to calculate the posterior analytically. For the next type of flight, we can then sample from this distribution to get the initial prior distribution for a, b again, and the cycle continues. Pseudo-code for the algorithm can be found in section B.

4 Results

4.1 Simulation model

To test the pricing model, a simulation engine has been created in order to test under different conditions. The simulation engine simulates a range of different types of flights, which all have the same type of distribution, but different parameters, for the number of passengers and the willingness-to-pay (WTP) per passenger. The simulation runs in two steps, first simulating flight data based on the input for the flight type. And second, for every flight within the flight type, using the earlier simulated data, drawing samples for this specific flight. As input for the simulation, the number of different flight types, the number of specific flights per type, and the number of simulations to run is configured. One round of the simulation model is defined as once going through all flight types, with all flights within that type. Going forward we compare results using 75 simulation rounds of 50 different flight types, with all 6 flights per type. This means that in every simulation round, there is a total of 300 flights ($50 \cdot 6$), and a full simulation of 75 rounds has a total of $22,500$ flights.

For every flight within a type of flight, the pricing model knows which type of flight is active, but has no further information about parameters for this type of flight. Before the start of the sales window in which seats are offered to all of the customers, the model has to set a price per seat. Then all customers are

in random order, as they are generated from the simulation, offered a seat and either buy a seat (if the seat price is lower or equal to their WTP) or not. This random order of customers would not be a valid assumption for regular ticket pricing. However for paid ancillary seats, this does seem applicable as these seats are often also sold during check-in. The model afterwards only learns the total number of seats that were sold, and how often a seat has been offered in total. See figure 5 for a visual representation of the interaction.

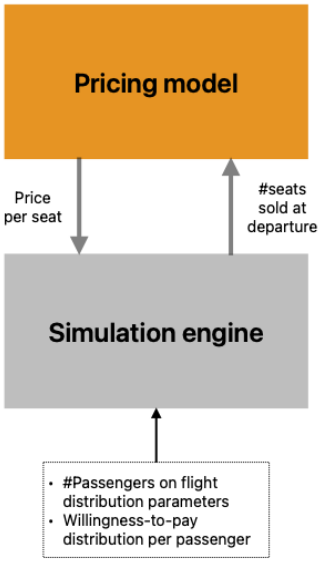


Figure 5: The interaction between the simulation engine and the pricing model for a specific flight.

As mentioned before, a customer will accept a seat if the price is lower than or equal to the WTP of this customer. However, there are a few further details. A customer is part of a booking, with a (random) number of other people in the booking. The seat is only accepted if the number of adjacent seats available, with an acceptable price, is at least as high as the number of people in the booking. Furthermore, from empirical results, it is known that for individual customers (the largest share of bookings), the middle seats are less attractive than window or aisle seats. So in this case, the simulation engine reduces the WTP to 25% of the initial WTP for this customer, for that specific middle seat. Using this simulation model, all different approaches described in section 3 are compared against each other. At the same time, a wide range of distributions is tested for the number of passengers that are on the flight, and the willingness-to-pay distribution of those customers.

All code for both the simulation model, and all models from section 3, is

available on *Simulation and optimization models for Bayesian Reinforcement Learning to Optimize Paid Ancillary Revenue in the Airline Industry paper* 2025.

4.2 Result comparison

There are two important factors to consider in this case. On the one hand, it is important that the algorithm 'quickly' learns, so that we don't lose a lot of revenue over hundreds/thousands of flights, before the model begins to perform well. On the other hand, that the model has a high maximum revenue. To compare across simulations how well an algorithm is performing, we compare against an algorithm that is all-knowing, so knows exactly which customers are on the flight and their exact willingness-to-pay. This algorithm can then sell separate seats at exactly the maximum willingness-to-pay to exactly the customers that are willing to pay the most. In order to measure both how quickly an algorithm is learning, and how well it is able to perform in the long run, we measure the percentage of theoretical maximum both for the first 3 simulation rounds (meaning, the first 900 flights optimized), and for the last 3 simulation rounds.

For the Q-Learning model of section 3.3 and the binomial model of section 3.7 it is necessary to upfront define a set of allowed prices, for all other models this is not necessary. In all following experiments, allowed prices were defined as prices between 1 and 250, in 100 equal steps.

The most important scenario is what we called the 'realistic setting'. In this case, we have tried to estimate distributions and parameters that appear to be close to results in reality. However, to our knowledge there are no public datasets available that can be used to estimate price elasticity from actual customer behavior, so we can't be certain that this is exactly right. Therefore, multiple other scenarios are also compared.

In a realistic setting in terms of distributions for the number of passengers and willingness-to-pay (meaning, giving similar results as observed at a large European airline as outcomes), the exponential model works very well, as can be seen in figure 6, where on the X-axis the number of simulation rounds is shown, and on the Y-axis the percentage of total revenue the model achieved on one full set of flights (so the 50 different types of flights, with all 6 flights per type, meaning in total on 300 flights) compared to the theoretical maximum from the all-knowing model. Since for every new round the number of passengers and WTP is sampled again, the model does learn, but can also have worse performance in a new round simply due to randomness. Dependent on the variability of the number of passengers and willingness to pay, the performance of the models gets closer to the theoretical maximum.

Testing across a wide range of different settings, the exponential model from section 3.8 always appears to work well. The model was tested against:

1. *Realistic settings*, where the simulation model is intended to give results similar to what seems to be relevant in practice:

Figure 6: Simulation results of the various models with realistic distribution.

   (a) *Number of passengers* is Binomially distributed with $n =$ max passengers, $p =$ probability per passenger to show up
      i. n is uniformly distributed between 80 and 120
      ii. p is uniformly distributed between 0.7 and 0.9
   (b) *Willingness-to-pay (WTP)* is log-normally distributed with $\mu = 0, \sigma$, multiplied by a scale parameter s
      i. μ is 0
      ii. σ is normally distributed with $\mu = 0.35, \sigma = 0.2$
      iii. s is Poisson-distributed with $\lambda = 40$

2. *High variability*, where the simulation model creates a situation with very high variability across parameters
   (a) *Number of passengers* is Binomially distributed with $n =$ max passengers, $p =$ probability per passenger to show up
      i. n is uniformly distributed between 50 and 150
      ii. p is uniformly distributed between 0.5 and 0.99
   (b) *Willingness-to-pay (WTP)* is Cauchy distributed with a μ and σ parameter that are both uniformly distributed
      i. $\mu \sim$uniform(40, 80)
      ii. $\sigma \sim$uniform(5, 20)

3. *Low variability*, where the simulation model creates a situation with a low variability across parameters

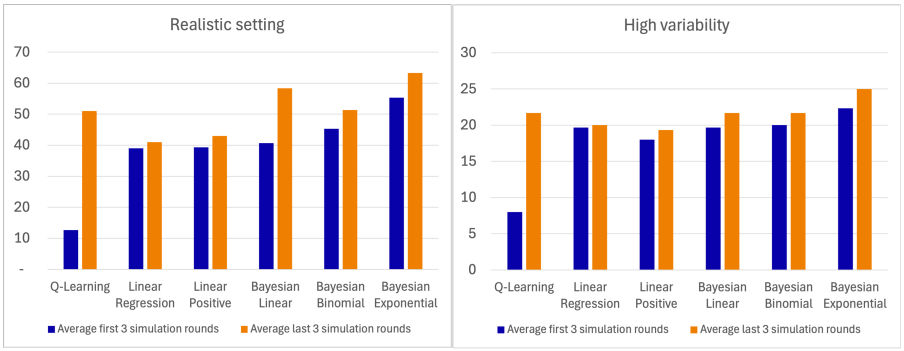
(a) *Number of passengers* is Binomially distributed with n = max passengers, p = probability per passenger to show up, both with fixed parameters
   i. n is 120
   ii. p is 0.9
(b) *Willingness-to-pay (WTP)* is normally distributed with a μ and σ parameter that are both fixed
   i. μ is 50
   ii. σ is 10

4. *Fixed #passengers, random WTP*, where the simulation model creates a situation with a fixed number of parameters, but a random willingness-to-pay
   (a) *Number of passengers* is fixed at 100
   (b) *Willingness-to-pay (WTP)* is log-normally distributed with $\mu = 0, \sigma$, multiplied by a scale parameter s
      i. μ is 0
      ii. σ is normally distributed with $\mu = 0.35, \sigma = 0.2$
      iii. s is Poisson-distributed with $\lambda = 40$

In all these scenarios, there is no relation or usage between the distribution used in the simulation, and the distributions used in the Bayesian approaches. So these are nowhere aligned to make sure there is a well-fitting prior for instance. See figure 7 for an overview of the results across these input-settings. The log-normal form for willingness-to-pay is closer to the exponential model from section 3.8, however, a logit-like shape is often used in practice and literature for willingness-to-pay, and various other distribution froms are tested as well. While we do often use an exponential-like functional form in the simulation, we also vary the underlying distributions (e.g., log-normal or Cauchy for WTP) to ensure that our Bayesian exponential approach is not artificially favored. Our experiments show that even when the simulation distribution departs from an exact exponential shape, the proposed model still learns effective pricing strategies.

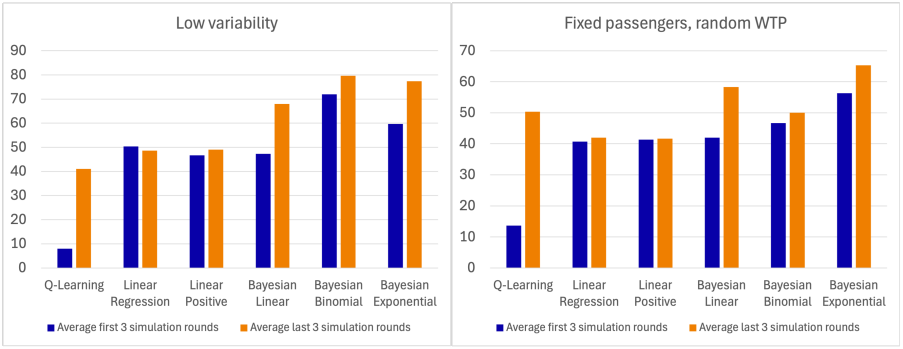
In general, it is clear that the exponential model is both learning quickly, as in the first 20 simulations it is commonly the best, or among the best, performing algorithms. And the same is the case at the last 25 simulations, where the model is also usually outperforming all other models. Another convenient property is that the standard deviation of the performance is very low compared to the other models, meaning that it is consistently performing at this level. Besides, in reality it is of course not possible to know what input distributions are most accurately reflecting reality. Another good property of the exponential model is that it is consistently performing well across scenarios, without the need to tune parameters for that specific scenario (across all tests, the hyper parameters - like prior-distributions - were kept the same). More detail about the sensitivity to the prior parameters can be seen in section A.3. From there it can be seen

that especially in the early stages, what prior is selected matters. But as long as the priors are 'wide' enough, in the end the models converge towards the same eventual performance. These values will in the end have to be estimated based on what data is available in practice.



(a) Simulation results with a realistic input distribution.

(b) Simulation results with a very high variability input distribution.



(c) Simulation results with a very low variability input distribution.

(d) Simulation results with a fixed number of passengers, and random willingness-to-pay.

Figure 7: Simulation results across various input distributions.

4.3 Computation times

The exponential model from section 3.8 can have relatively large computation time compared to the other methods due to the sampling involved. We have found on a regular laptop (Apple with an Apple M1 Pro chip), that using the described configuration, a full simulation of 75 rounds, which contains 22,500 flights, takes about 7,800 seconds. This means that about 2.9 flights per second can be optimized using this method. This is much heavier than the other methods, which all take between 15 and 20 seconds to run for a full simulation – meaning about 1,285 flights per second. However, since the model only needs

to run once after a flight departs, this computational overhead is feasible for day-to-day airline operations.

Taking the main numbers from OAG 2023 which states that on average around 27% of ancillary revenues for full-service carriers comes from seat selection and upselling. From the same report, total ancillary revenue for large network carriers can be over one billion USD, and for some large carriers over 5 billion USD (for American, Delta and United). This means that seat ancillary revenue for those carriers is over 270 million USD, up to 1.35 billion USD. The 4 to 5% percentage performance increase of the Bayesian exponential model over models with lower computation times with that is normally worth it.

5 Conclusion

We have shown that by using a model that uses a limited number of parameters, we can quickly learn the right settings and optimize paid seat prices without a lot of historical data. In almost all contexts on demand distributions, the exponential model from section 3.8 outperforms other approaches. The model performs well both against model-free approaches that can learn any relation between price and number of seats sold, and against other simple models that require less computation. Besides, the model is easy to implement in practice, as the data required is very limited. Only the price that the model itself outputs, and the number of seats that are sold at this price-level at departure are needed. Both are readily available in any practical context. Next to that, the model requires feasible computational power. Prices are set for the duration of the flight sales window, and learning - which is computationally the most expensive - is only done after a flight departs. In any practical context, this can easily be handled computationally.

Future research that would be of interest is in three directions. First, more parameters could be used to estimate the optimal price point. For instance, it would be relevant to use more parameters about the flight itself such as flight duration, day or night flight, etc. Also more customer parameters could be incorporated, to investigate an approach that combines a customer choice type of approach, with this revenue management type of approach. A very interesting direction would be to combine a customer choice model approach, which is most often found in literature, with a revenue management approach also taking into account the inventory like in our paper. It's an open question whether this revenue management approach works better than a customer choice type approach. Second, price determination can be more dynamic than it is in this model. Instead of fixing the price across seats and during the 'lifetime' of the flight, this could also be varied both across specific seats (e.g. it is generally known that window or aisle seats are more attractive than middle seats), and while seats are being sold to optimize pricing based on intermediate feedback. Third, we still use a heuristic here in sections 3.7 and 3.8 to optimally choose a price. We could potentially also model the system as a Markov Decision Process (MDP) and then use value iteration to find the optimal solution.

Funding and Competing Interests

This research was supported by the institutions affiliated with the authors, with no external funding received. The authors declare that they have no competing interests.

References

Babić, Ružica Škurla, Maja Ozmec Ban, and Jasmin Bajić. 2019. "A review of recent trends in airline ancillary revenues." *Emc Review-Economy and Market Communication Review* 17 (1).

Bastani, Hamsa, David Simchi-Levi, and Ruihao Zhu. 2022. "Meta dynamic pricing: Transfer learning across experiments." *Management Science* 68 (3): 1865–1881.

Belobaba, Peter P. 1989. "OR practice—application of a probabilistic decision model to airline seat inventory control." *Operations research* 37 (2): 183–197.

Boer, Arnoud V den, and Bert Zwart. 2014. "Simultaneously learning and optimizing using controlled variance pricing." *Management science* 60 (3): 770–783.

Chib, Siddhartha, and Edward Greenberg. 1995. "Understanding the metropolis-hastings algorithm." *The American Statistician* 49 (4): 327–335.

Den Boer, Arnoud V. 2015. "Dynamic pricing and learning: historical origins, current research, and new directions." *Surveys in operations research and management science* 20 (1): 1–18.

Dogan, Ibrahim, and Ali R Güner. 2015. "A reinforcement learning approach to competitive ordering and pricing problem." *Expert Systems* 32 (1): 39–48.

Fink, Daniel. 1997. "A compendium of conjugate priors." *Tech. Rep.* 46.

Gosavi, Abhijit. 2004. "A reinforcement learning algorithm based on policy iteration for average reward: Empirical results with yield management and convergence analysis." *Machine Learning* 55:5–29.

Harrison, J Michael, N Bora Keskin, and Assaf Zeevi. 2012. "Bayesian dynamic pricing policies: Learning and earning under a binary prior distribution." *Management Science* 58 (3): 570–586.

Hastings, W Keith. 1970. *Monte Carlo sampling methods using Markov chains and their applications.* Oxford University Press.

Hoffman, Matthew D, Andrew Gelman, et al. 2014. "The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo." *J. Mach. Learn. Res.* 15 (1): 1593–1623.

Kummara, MadhuSudan Rao, Bhaskara Rao Guntreddy, Ines Garcia Vega, and Yun Hsuan Tai. 2021. "Dynamic pricing of ancillaries using machine learning: one step closer to full offer optimization." *Journal of Revenue and Pricing Management* 20 (6): 646–653.

Luo, Yiyun, Will Wei Sun, and Yufeng Liu. 2024. "Distribution-free contextual dynamic pricing." *Mathematics of Operations Research* 49 (1): 599–618.

Mishra, Pankaj, Ahmed Mosutafa, and Takayuki Ito. 2020. "Reinforcement Learning based Real-Time Reserve Price Optimisation in Dynamic Environments." In *The 23rd International Conference on Principles and Practice of Multi-Agent Systems.*

Mumbower, Stacey, Susan Hotle, and Laurie A Garrow. 2022. "Highly debated but still unbundled: The evolution of US airline ancillary products and pricing strategies." *Journal of Revenue and Pricing Management,* 1–18.

OAG. 2023. "Shaping Airline Retail: The Unstoppable Rise of Ancillaries." Accessed October 27, 2024. https://www.oag.com/blog/shaping-airline-retail-unstoppable-rise-ancillaries.

Ødegaard, Fredrik, and John G Wilson. 2016. "Dynamic pricing of primary products and ancillary services." *European Journal of Operational Research* 251 (2): 586–599.

Otero, Daniel F, and Raha Akhavan-Tabatabaei. 2015. "A stochastic dynamic pricing model for the multiclass problems in the airline industry." *European Journal of Operational Research* 242 (1): 188–200.

Ren, Xinhui, Na Pan, and Hong Jiang. 2022. "Differentiated pricing for airline ancillary services considering passenger choice behavior heterogeneity and willingness to pay." *Transport Policy* 126:292–305.

Shao, Shuai, and Göran Kauermann. 2020. "Understanding price elasticity for airline ancillary services." *Journal of Revenue and Pricing Management* 19:74–82.

Shukla, Naman, Arinbjörn Kolbeinsson, Ken Otwell, Lavanya Marla, and Kartik Yellepeddi. 2019. "Dynamic pricing for airline ancillaries with customer context." In *Proceedings of the 25th ACM SIGKDD International Conference on knowledge discovery & data mining,* 2174–2182.

Simulation and optimization models for Bayesian Reinforcement Learning to Optimize Paid Ancillary Revenue in the Airline Industry paper. 2025. https://osf.io/bk8tc/?view_only=eb7eb8a0700c4f2ca1788457c675609a. Anonymized for peer review.

Thompson, William R. 1933. "On the likelihood that one unknown probability exceeds another in view of the evidence of two samples." *Biometrika* 25:285–294.

Wang, Kevin K, Michael D Wittman, and Adam Bockelie. 2021. "Dynamic offer generation in airline revenue management." *Journal of Revenue and Pricing Management* 20:654–668.

Warnock-Smith, David, John F O'Connell, and Mahnaz Maleki. 2017. "An analysis of ongoing trends in airline ancillary revenues." *Journal of Air Transport Management* 64:42–54.

Wilson, John G, and Feryaal F Ahmed. 2024. "Pricing Optimization for Perishable Primary and Ancillary Items with Time-Dependent Inventory Drawdown." *IFAC-PapersOnLine* 58 (19): 592–597.

Wittman, Michael D, and Peter P Belobaba. 2019. "Dynamic pricing mechanisms for the airline industry: A definitional framework." *Journal of Revenue and Pricing Management* 18:100–106.

Yang, Yang, Wan-Ling Chu, and Cheng-Hung Wu. 2022. "Learning customer preferences and dynamic pricing for perishable products." *Computers & Industrial Engineering* 171:108440.

Zhao, Guihong, Yue Cui, and Shaoyu Cheng. 2021. "Dynamic pricing of ancillary services based on passenger choice behavior." *Journal of Air Transport Management* 94:102058.

A Sensitivity of parameters

A.1 Sensitivity of Q-Learning parameters

Q-Learning can be sensitive to parameter changes in α - the learning rate, γ the discount factor, and ϵ, the parameter that governs the amount of exploration vs exploitation. Below table gives an overview of the impact of the parameters, using the default simulation scenario from section 4.2. The optimal parameters from this overview are used in the final results comparison in that section.

	Performance first 3 rounds	Performance last 3 rounds
$\alpha = 0.85$	12%	38%
$\alpha = 0.92$	12%	51%
$\alpha = \mathbf{0.995}$	**13%**	**51%**
$\gamma = 0.85$	14%	38%
$\gamma = \mathbf{0.9}$	**13%**	**51%**
$\gamma = 0.95$	8%	38%
$\epsilon = 0.001$	1%	28%
$\epsilon = \mathbf{0.01}$	**13%**	**51%**
$\epsilon = 0.05$	45%	49%

A.2 Sensitivity of linear regression parameters

Linear regression can be sensitive to parameter changes in ϵ, the parameter that governs the amount of exploration vs exploitation. Below table gives an

overview of the impact of this parameter, using the default simulation scenario from section 4.2. The optimal parameter from this overview is used in the final results comparison in that section.

	Performance first 3 rounds	Performance last 3 rounds
$\epsilon = 0.01$	41%	43%
$\epsilon = \mathbf{0.05}$	**41%**	**43%**
$\epsilon = 0.1$	39%	43%

A.3 Sensitivity of prior-parameters for exponential model

The exponential model has various prior parameters which can influence the performance of the algorithm, especially in the initial rounds of simulations and learning. We have not fine-tuned the prior parameters for the model, since that would also not be possible in practice – as in practice you never observe the percentage of theoretical maximum. However, we did select prior parameters that seemed logical.

Five scenarios were run against the same simulation model, using the realistic settings for number of passengers and willingness-to-pay:

	intercept μ	slope μ
Used parameters	6.0	-0.09
Intercept μ increased	9.0	-0.09
Intercept μ decreased	3.0	-0.09
Slope μ increased	6.0	-0.01
Slope μ decreased	6.0	-0.18

Results for these scenarios can be seen in figure 8, from which it's visible that especially in the initial rounds differences in performance are larger. In later rounds, these differences converge towards the same performance. Since in practice theoretical maximum performance is not observable, and the priors likely differ per situation, it is not possible to give a concrete recommendations for these values, and they would depend on what kind of information is available per situation to estimate them.

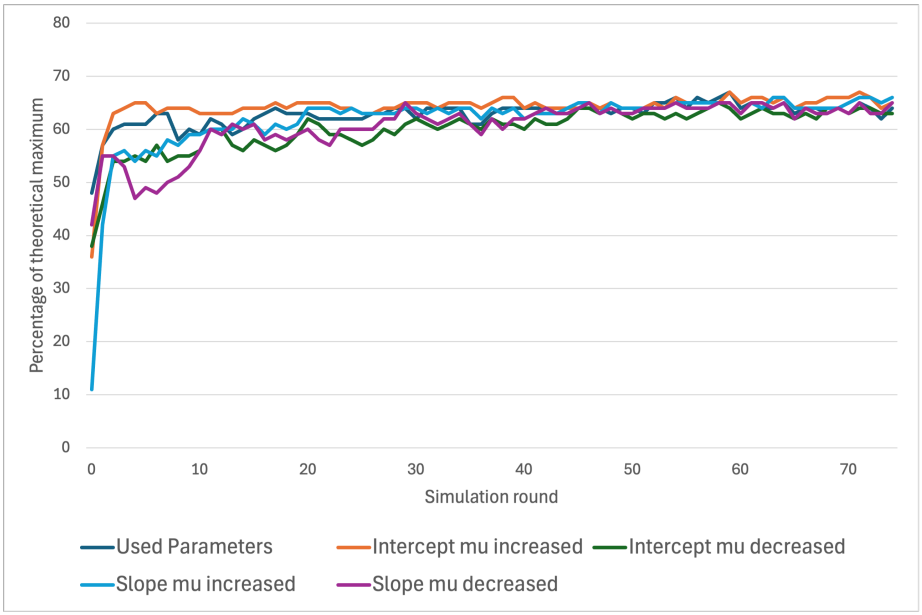


Figure 8: Simulation results showing the sensitivity to the prior parameters for the exponential model.

B Pseudo-code for the exponential model

Algorithm 1 Exponential Pricing Model with Bayesian Updating

Initialization:

- Define hierarchical priors for intercept and slope across N flight types.
- Define prior distributions
- Initialize sampling parameters (e.g., MCMC samples, burn-in, lag).

Main Algorithm:

1. **Initialize Flight Type Parameters:**
   - For each flight type, draw initial values of slope and intercept from hierarchical priors based on Thompson-sampling.
2. **Compute Prediction for a Given Price:**
   - Use posterior intercept μ_b and slope μ_a to compute expected demand:

$$\hat{D} = \frac{C}{1 + e^{-(\mu_b + \mu_a \cdot \mathcal{P})}}$$

3. **Select Optimal Price:**
   - For each price $p \in \mathcal{P}$:
     - Compute revenue:

$$\text{Revenue}(p) = \hat{D}(p) \cdot p$$

   - Select price p^* that maximizes revenue.
4. **Update Model Post-Flight:**
   - Observe total seats sold s and update posterior distributions through MH-sampling:

$$\text{Posterior}(\mu_b, \mu_a) \propto \text{Likelihood}(\text{Data}|\mu_b, \mu_a) \cdot \text{Prior}(\mu_b, \mu_a)$$

   - Update hierarchical priors based on observed distributions.
5. **Repeat:**
   - Use updated priors and posteriors for subsequent flights of the same type, again using Thompson-sampling to get actual values as in step 1.