Revenue Management under Reviews and Online Ratings

D.D. Sierag[a,b,*], D.M. Roijers[c,d]

[a] *CWI, Stochastics Department, Amsterdam, The Netherlands*

[b] *VU University Amsterdam, Faculty of Exact Sciences, Amsterdam, The Netherlands*

[c] *Vrije Universiteit Brussel, Department of Computer Science, Brussels, Belgium*

[d] *University of Oxford, Department of Computer Science, Oxford, United Kingdom*

Abstract

This article proposes a revenue management model that integrates reviews and ratings. The dependency between reviews and revenue is two-fold: the content of a review depends on the product the customer purchases, and reviews impact the demand. A complicating factor in this model is that the effects of reviews are delayed, i.e., by sacrificing revenue now in order to get better reviews, long-term revenue can be increased. Because the full planning problem of finding an optimal strategy for the proposed model is intractable, a novel solution methodology is proposed to solve the problem approximately by restricting the space of possible solutions to *target review strategies*. It is shown that target review strategies for the full problem can be found by viewing the full problem as a series of multi-objective Markov decision processes subproblems, and aiming for a target review ratio in the subproblems while optimising revenue. Numerical studies show that taking reviews into account in this manner can lead to an increase in revenue of up to 11% compared to the case where the sole objective is revenue.

Keywords: Multi-Objective Markov Decision Processes, Revenue Management, Online ratings, reviews

1. Introduction

A recent development that influences sales in various industries is the wide availability of reviews. For example, customers who attended a play in a theatre are given the opportunity to share their experience with other potential clients via websites like Tripadvisor. Research strongly suggests that reviews influence the buying behaviour of customers [9, 10, 26]. For instance, [18] show that 'consumers seem to be more influenced by early negative information' and [11, 22, 25] show that more recent reviews have more impact. It is therefore in the interest of the theatre to get good reviews and avoid bad reviews. Also, the other way around, customer reviews are influenced by the price/quality perception of the customer

*Corresponding author. Postal address: Science Park 123, 1098 XG, Amsterdam, The Netherlands. Tel.: +31(0)20 592 4168. Email address: dirk@cwi.nl

[27]. This article presents the first study that proposes a decision-theoretic model that explicitly includes reviews to maximise the revenue.

This article studies the problem of optimising revenue over multiple performances when demand depends on reviews, and where cancellations, overbooking, and customer choice preferences are taken into account. Moreover, the model contains a feedback mechanism: on one hand the creation of reviews depends on the purchases, and on the other hand demand depends on reviews. The focus is on recent reviews instead of rating, since rating tends to remain more or less constant, while customers are more influenced by recent reviews, either positive or negative. Demand for a certain performance depends on reservations of previous performances, due to the reviews that are released in the meantime. This makes it a hard problem, suffering heavily from the curse of dimensionality. The approach of this article is to maximise revenue while the ratio between positive and negative reviews remains constant, called *target review strategies*[1]. Searching in the space of target review strategies is an approximation, yet even under these circumstances finding an optimal solution is challenging. An elaborate though tractable solution method is provided, using a multi-objective Markov decision process formulation.

This article builds on three research areas. The first research area is the work on choice-based assortment problems, such as the work by [17, 19] (see [16] for an overview of literature in this area). The basic idea is that customer preferences depend on the products that are offered, and the set of offered products should be optimised accordingly. The second area is the work on e-word-of-mouth (eWOM), in particular the effect of reviews on demand, and the effect of purchases on reviews. See for example the survey by [25] for an overview of related literature. Studies often show that the impact of reviews is strong, though no connection is made to optimise assortments accordingly. Closest to the work in this article is the article by [4], where reviews are used to forecast box office revenues. However, the reviews are not used to optimise revenue, and the creation of new reviews is not considered. The third research area is the work on multi-objective Markov decision processes [14]. When it is not clear a priori what the relative importance of each objective is and how the objectives interact, a set of different trade-offs between the objectives is considered. Such a set is called a coverage set. In this article, a series of multi-objective problems approximates an otherwise intractable problem.

This article makes the following research contributions: 1) a model that incorporates reviews in the optimisation process (Section 2); 2) an analysis of the model indicating that the full problem is intractable, and an approximation as series of tractable multi-objective problems (Sections 3 and 4); and 3) numerical experiments indicate that taking reviews into account induces significantly higher long-term revenue

[1]Note that this is not the only possible solution approach. However,this solution has the advantage that it is tractable and leads to large improvements in revenue as Section 5 indicates.

than when solely revenue is optimised (Section 5). Finally, implications for research and practice are discussed in Section 6.

2. General Model Description

The model is framed in the context of theatres. However, the model is more general and can be applied to other single-leg experience-based products, like sports events or concerts, or the cinema. In general, the approach can be used for any problem that can be modelled as a series of decision problems in which the output of an earlier problem (partially) parameterises the next problem in the series.

Purchasing behaviour of theatre clients is influenced by reviews, and, on the other hand, the *price* (and conditions) of the ticket influence customers that write reviews. By offering a lower price, for example, the theatre can attract different market segments. Clients with a low budget might appreciate the play better than high-end customers, and can lead to better reviews. By deciding which products to offer at what time, depending on demand and availability, the theatre can control revenues as well as reviews.

In practice, theatres have multiple seating areas, similar to airlines and hotels, which have different seating classes and room types. However, considering multiple seating classes complicates the structure of the problem and distracts from the focus and core findings of this paper. Therefore, in line with literature on hotel and airline RM [8, 17, 19], this paper studies the identical-seat single-leg approach to focus on the challenges of reviews rather than the challenges of multiple seating areas.

Consider a theatre with C identical seats that wants to sell them for one or more plays on multiple performances. In particular, assume a (possibly infinite) set $I \subset \mathbb{N}$ of future arrival dates, each with booking horizon of T time units (selling seats of a particular performance only starts at T units before the performance), 0 being the arrival time. Overbooking is allowed up to $C_{\max}$ seats.[2] Every customer has the opportunity to give a *review* of their experience, which is *positive* and/or *negative*, and corresponds to a performance $i \in I$ of the past. The set of past reviews is given by $\mathcal{R}$.

Demand depends on the reviews. Clients for performance $i \in I$ arrive according to a Poisson process $\lambda_i(\mathcal{R})$ dependent on $\mathcal{R}$, starting T time units before performance i. Each seat is sold as a *fare product* j, which is a combination of a seat with a price r_j and conditions, such as the cancellation policy. Moreover, customers are influenced by the fare product when writing a review. The probability that fare product j leads to a positive review is q_j^p and to a negative review it is q_j^n. The probability of writing a negative or positive review is independent, so a customer can write a review that has both positive and negative

[2] NB, $C_{\max}$ can be equal to C, hence the model without overbooking is a special case of this model.

components with probability $q_j^p q_j^n$. It is also possible the client writes no review at all with probability $1 - q_j^p - q_j^n - q_j^p q_j^n$. It is assumed that the probabilities are well defined, i.e., $q_j^p + q_j^n + q_j^p q_j^n \leq 1$.

Assume there is a finite number of fare products $N = \{1, \ldots, n\}$. At each moment in time, for each future performance separately that is at least T time units ahead, the theatre manager decides which *offer set* $S \subset N$ of fare products to offer. Depending on the offer set S displayed and the reviews $\mathcal{R}$, an arriving customer decides to either buy one of the fare products $j \in S$, with probability $P_j(S)$, or leave and buy nothing at all, with probability $P_0(S)$.

Reservations are allowed to be cancelled, where a potential refund depends on the cancellation conditions of the product. Assume that cancellations happen independently from each other. The cancellation rate $\gamma_i(\mathcal{R})$ is independent of the product, but depends on the reviews. If there are x_j^i reservations for fare product j for performance i, then cancellations occur with rate $\gamma_i(\mathcal{R}) x_j^i$. The costs for the theatre of a cancelled reservation of product j at time t equals $c_j^i(t)$. When there are more reservations than capacity at the end of the performance, i.e., when $x_i > C_{\max}$, where $x_i = \sum_{j \in N} x_j^i$, a penalty $q(x_i)$ is incurred. See Figure 1 for an overview of the model.

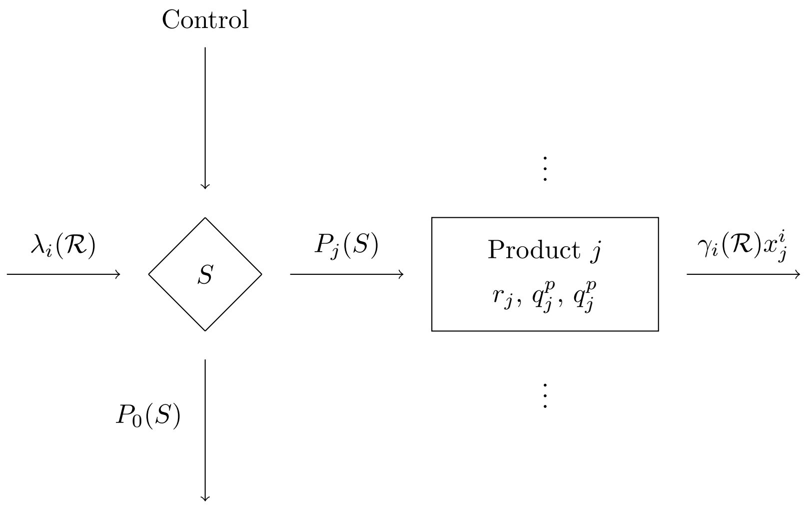

Figure 1: Visualisation of the model. Per performance the arrival process is Poisson distributed with parameter $\lambda_i(\mathcal{R})$. The manager controls the offer set S. Under this offer set an arriving customer buys product $j \in S$ with probability $P_j(S)$. With probability $P_0(S)$ the customer buys nothing. Finally, cancellations of product j follow an exponential distribution with parameter $\gamma_i(\mathcal{R})$.

2.1. Modelling Demand and Cancellations as a Function of Reviews

Demand is influenced by reviews, either positive or negative [9, 10, 18, 22, 26]. To capture this effect, consider the following set-up. Let Q_i^p and Q_i^n be the number of positive and negative reviews resulting from performance i. Recent reviews are more relevant than older reviews. Let $\alpha \in (0, 1)$ be the discounting parameter for relevance of reviews and let M be the number of past performances that are relevant.

Define the *discounted reviews* $\tilde{Q}_k^p$ and $\tilde{Q}_k^n$ on performance k by

$$\tilde{Q}_k^p := \sum_{i=k-M}^{k-1} \alpha^{k-i-1} Q_i^p, \qquad \tilde{Q}_k^n := \sum_{i=k-M}^{k-1} \alpha^{k-i-1} Q_i^n.$$

The most recent reviews of performance $k-1$ are not discounted, while the relevant reviews of arrival $k-M$ are discounted most, by factor α^{M-1}.

Let $\bar{\lambda}_i, \bar{\gamma}_i \in \mathbb{R}$ be the base parameter of demand and cancellations, respectively, for performance $i \in I$. Let $\beta_p^\lambda, \beta_n^\lambda, \beta_p^\gamma, \beta_n^\gamma \in \mathbb{R}$. Define $\lambda_i(\mathcal{R})$ and $\gamma_i(\mathcal{R})$ by

$$\lambda_i(\mathcal{R}) := \bar{\lambda}_i \exp\left(\beta_p^\lambda \tilde{Q}_k^p + \beta_n^\lambda \tilde{Q}_k^n\right),$$
$$\gamma_i(\mathcal{R}) := \bar{\gamma}_i \exp\left(\beta_p^\gamma \tilde{Q}_k^p + \beta_n^\gamma \tilde{Q}_k^n\right).$$

In accordance with literature [22] it is assumed that positive reviews have positive effect on demand and negative reviews have a negative effect on demand, which translates to $\beta_p^\lambda > 0$ and $\beta_n^\lambda < 0$. For cancellations the opposite is expected: positive reviews result in less cancellations and negative reviews result in more cancellations, i.e., $\beta_p^\gamma < 0$ and $\beta_n^\gamma > 0$. In line with findings in literature [18], the effect of negative reviews is greater than positive reviews: $\beta_p^\lambda < -\beta_n^\lambda$ for demand, and $\beta_p^\gamma < -\beta_n^\gamma$ for cancellations. Furthermore, we expect the arrival rate to correspond to rankings in search results that are heavily influenced by these reviews; the better the reviews, the better the ranking in the list. The ranking is then presented to a user, and the arrival rate corresponds to the arrival rate at the list times the probability of a person clicking on the item in the list corresponding to our booking process. Because it is a well-known phenomenon in search engines that the probability of a click decays rapidly with the position in a search-result list [7][3] The exponential function provides these intuitive features, as well as the fact that demand cannot be negative.

Define the *review ratio* ρ_k by

$$\rho_k := \frac{\tilde{Q}_k^p}{\tilde{Q}_k^p + \tilde{Q}_k^n}.$$

In the remainder of this paper, write ρ instead of ρ_k for notational convenience. Then an equivalent

[3]Besides user studies on a smaller scale, like the seminal paper by Joachims et al. [7], such results have been duplicated by commercial parties, some of which has been made available online as well. See e.g., https://www.sistrix.com/blog/click-probabilities-in-the-google-serps/. — a fact exploited in learning-to-rank algorithms for search engine optimisation [5] — we want to have a function that expresses this decay.

formulation for $\lambda_i(\mathcal{R})$ and $\gamma_i(\mathcal{R})$ is

$$\lambda_i(\mathcal{R}) := \bar{\lambda}_i \exp\left(\tilde{\beta}_p^\lambda \rho + \tilde{\beta}_n^\lambda (1-\rho)\right),$$
$$\gamma_i(\mathcal{R}) := \bar{\gamma}_i \exp\left(\tilde{\beta}_p^\gamma \rho + \tilde{\beta}_n^\gamma (1-\rho)\right),$$

where $\tilde{\beta}_p^\lambda = \beta_p^\lambda(\tilde{Q}_k^p + \tilde{Q}_k^n)$, $\tilde{\beta}_n^\lambda = \beta_n^\lambda(\tilde{Q}_k^p + \tilde{Q}_k^n)$, $\tilde{\beta}_p^\gamma = \beta_p^\lambda(\tilde{Q}_k^p + \tilde{Q}_k^n)$, and $\tilde{\beta}_n^\gamma = \beta_n^\gamma(\tilde{Q}_k^p + \tilde{Q}_k^n)$.

2.2. Running Example

To illustrate the model, consider the following running example. To capture the effects of reviews under different circumstances, four instances are considered by adjusting the review probabilities and the review parameters of the demand. For both cases the effects are either large or small, leading to four scenarios: 1) large effect on both demand and review probabilities; 2) large effect on demand and small effect on review probabilities; 3) small effect on demand and large effect on review probabilities; 4) small effect on both demand and review probabilities.

Part of the parameters that are used are based on the example used in the numerical results section of [17] and [19]. Let $n = 10$ be the number of products sold with corresponding price vector

$$r = (240, 220, 190, 160, 120, 112, 96, 80, 74, 70).$$

Overbooking is allowed up to 20% of the total capacity $C = 200$. Time is discretised to $T = 1000$. The demand, cancellation rate, and purchase probabilities are independent from the time period t. The base demand per time unit is equal to $\bar{\lambda}_i = 0.2$, for all performances $i \in \mathcal{I}$. The cancellation rate is assumed to be $\gamma = 0.0004$. Furthermore, define $(\beta_p^\lambda, \beta_n^\lambda) = (1, -1.5)$, i.e., a large effect on demand, in case 1) and 2); and $(\beta_p^\lambda, \beta_n^\lambda) = (0.9, -1)$, i.e., a small effect on demand, in case 3) and 4). The review probabilities are presented in Table 1 below. A difference is made between a reservation that is overbooked and one that is not. The difference is independent of the product, such that it can be incorporated in a similar way that overbooking costs are incurred. The values of the attributes are motivated by the findings of [27].

In this example, the purchase probabilities are modelled according to the commonly used *multinomial logit model (MNL)*, where the sole attribute is price. The results of this paper are more general, and the MNL model is only used as an illustration. Utility plays a crucial role in the MNL model [2]. Let $u_j = \beta r_j$ be the mean utility of product j, where $\beta \in \mathbb{R}$ is the corresponding weight; and let u_0 be the

	Large effect (cases 1 and 3)				Small effect (cases 2 and 4)			
	No overbooking		Overbooking		No overbooking		Overbooking	
Product j	q_j^p	q_j^n	q_j^p	q_j^n	q_j^p	q_j^n	q_j^p	q_j^n
1	0.160	0.120	0.140	0.140	0.130	0.098	0.110	0.118
2	0.130	0.160	0.110	0.180	0.090	0.144	0.070	0.164
3	0.210	0.095	0.190	0.115	0.155	0.088	0.135	0.108
4	0.240	0.080	0.220	0.100	0.170	0.082	0.150	0.102
5	0.230	0.110	0.210	0.130	0.140	0.124	0.120	0.144
6	0.288	0.056	0.268	0.076	0.194	0.072	0.174	0.092
7	0.304	0.048	0.284	0.068	0.202	0.069	0.182	0.089
8	0.270	0.090	0.250	0.110	0.160	0.116	0.140	0.136
9	0.326	0.037	0.306	0.057	0.213	0.065	0.193	0.085
10	0.280	0.085	0.260	0.105	0.165	0.114	0.145	0.134

Table 1: Review probabilities for both large and small effect. In case of overbooking the probabilities differ.

no-purchase utility. The purchase probabilities are then given by

$$P_j(S) = \begin{cases} \dfrac{e^{u_j}}{\sum_{i \in S} e^{u_i} + e^{u_0}} & \text{if } j \in S, \\ 0 & \text{otherwise,} \end{cases}$$

for all $j \in N$. Assume there are two types of customers: high price-sensitive and low price-sensitive. High price-sensitive customers have parameter $\beta^H = -0.005$, while low price-sensitive customers have parameter $\beta^L = -0.0015$.

3. A Tractable Model

This section formulates the optimisation problem of maximising revenue in the long run by taking into account the reviews. First, computational issues related to the size of the problem are discussed. Second, after concluding that the full problem is intractable, a model that approximates the large problem by breaking it down to a series of subsequently smaller problems is proposed. Within this (infinite) series, the proposed solutions are those that lead to stable review sets. Such solutions are referred to as *target review strategies*. On the level of a smaller problem, the review sets must be stable for a target review strategy, for two objectives: positive reviews, and negative reviews. These objectives supplement the (immediate) revenue objective. A dynamic programming formulation is given to find the resulting three-dimensional value function. In the long run, the review scores influence the long-term revenue. However, it is not clear at the level of a smaller problem how the review scores translate to long-term effects on the revenue. Therefore, in this section the focus lays on a three-objective model that incorporates the different trade-offs between these objectives. In Section 4 the computational method to find target review strategies is discussed.

3.1. Computational Issues

Suppose that the sales manager needs to optimise revenue for multiple shows or performances. Then the creation of new reviews after a show influences the purchasing behaviour of customers in the future. Moreover, the reviews may be updated during the booking horizon of another performance. Hence, strategies for a particular performance need to take into account the strategies of other performances.

To illustrate this, consider the following example, with only three performances with a booking horizon of three days. Only the reviews from the last performance are used, i.e., $M = 1$. Denote S_t^i as the offer set that is offered t days beforehand for performance i. The offer sets of the first performance are not influenced by any previous decisions, as well as the first two offer sets of performance 2 and the first of performance 3. However, after the first performance, reviews are updated. The number of positive and negative reviews depends on offer sets S_1^1, S_2^1, and S_3^1. Therefore, the decision which offer set S_1^2 to use on the last day before performance 2 depends on all offer sets of performance 1. Note that S_2^3 is also dependent on all offer sets of performance 1. Furthermore, after performance 2 the reviews are updated, such that S_1^3 depends on all offer sets of arrival 2, and hence also on all offer sets of performance 1. See Figure 2 for an overview of this example.

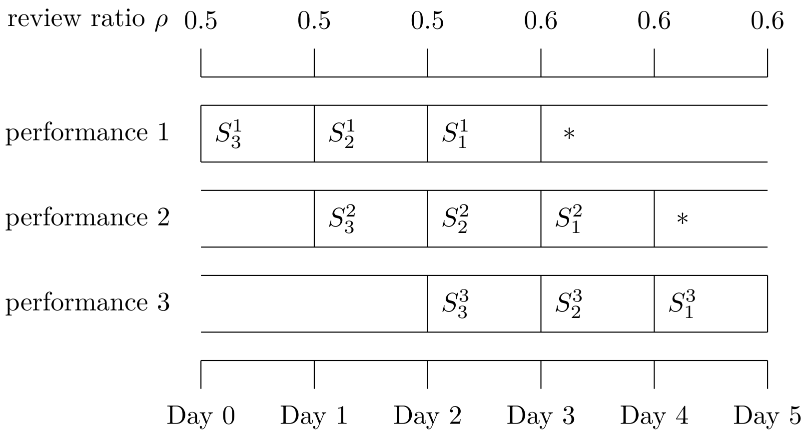

Figure 2: Visualisation of example with $M = 1$, $K = 0$, and $T = 3$ over a five day period. A star $*$ denotes the end of a performance and the update of the reviews.

Let R^i be the revenue that results from performance i and denote by $S^i = \{S_T^i, \ldots, S_1^i\}$ the strategy of performance i. Then R^i is a function of the strategies of all performances up to day i: $R^i\big(S^1, \ldots, S^i\big)$. The objective $\phi^I(S)$ to maximise the finite horizon problem, with $I \in \mathbb{N}$ the horizon length, is therefore given by

$$\phi^I(S) = \sum_{i=1}^{I} R^i\big(S^1, \ldots, S^i\big),$$

while the a-discounted infinite horizon objective function $\phi^{\infty}(S)$ is given by

$$\phi^{\infty}(S) = \sum_{i=1}^{\infty} a^{i-1} R^i \big(S^1, \dots, S^i\big).$$

At every moment in time products for up to T performances are in consideration to be for sale. Therefore, the state space for these T parallel processes has to be taken into account when solving the problem. Just using the Cartesian product of the T processes is not enough. Namely, this leads to a system that does not have the Markov property, because to make a decision it is necessary to take the review ratio ρ into account. One option is to include ρ in the state space as a real number. However, this leads to a continuous state-space, yielding an intractable problem. An option with finite state space is to keep track of the strategies of all past performances that are necessary for calculating ρ. This model has the Markov property. However, for both the finite horizon and the discounted infinite horizon problem the action space as well as the state space are still too large: the action space has size N^T and the state space has size up to N^I for the finite horizon, and converges to an infinite state space for the infinite horizon problem. Therefore, finding an optimal solution in the set of all feasible solutions is intractable because of the curse of dimensionality.

3.2. Target Review Strategies

Motivated by these deliberations, consider the set of *target review strategies*, which consists of strategies that keep the expected review ratio ρ constant and equal to a *target review ratio* ρ^*. The theatre can select ρ^* beforehand, and then maximise revenues under the condition that $\rho = \rho^*$ by deciding which products S_i to offer. What value of ρ^* the theatre should choose, the one that is most profitable, will be derived in the remainder of this paper.

When the target review ratio is achieved, i.e., when $\rho = \rho^*$, and remains fixed (in expectation), it is clear beforehand that the expected review ratios ρ at all time periods will be equal to ρ^* for each performance. Also, the long-term revenue is no longer influenced by changing review ratios; R_i solely depends on S_i. Therefore, each target review strategy can be solved separately and the problem can be solved per target review ratio. In such cases, the long term revenue is equal to the number of times a *single-performance dynamics problem* is executed, times the revenue attained in this single-performance dynamics problem, given the fixed target review ratio. Such a single-performance dynamics problem is modelled as a multi-objective Markov decision process and is tractable.

However, before the system reaches ρ^*, it must be show that ρ^* can in fact be reached. In order to do so, the process of Figure 3 is followed. An initial MOMDP is constructed on the basis of initial review scores, i.e., the parameters for demand and review probabilities are calculated from the past reviews. Then,

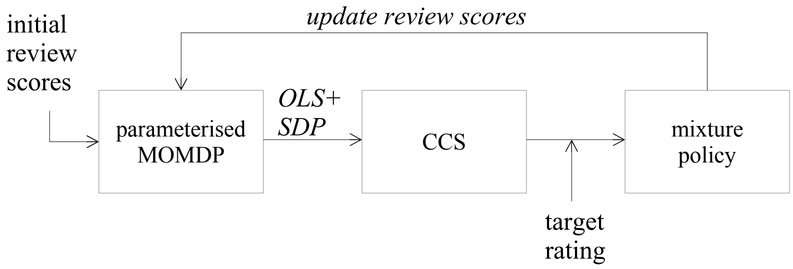

Figure 3: The process of finding a target review policy. Starting from an initial parameterised MOMDP, a CCS is identified, from which the mixture policy with the intended target review ratio is chosen, of which the output in terms of positive and negative review scores are used reparameterise the MOMDP. This process repeats until the parameters of the MOMDP converge.

a *convex coverage set*, which is formally defined in Section 4.1, is computed, from which a stochastic mixture policy can be constructed that optimises the short-term revenue while making the review scores converge to ρ^*. Executing this stochastic mixture policy leads to new review scores. With these new review scores, the process is restarted until the parameters of the MOMDP converge. After that, the maximal revenue for the target review ratio is determined, as the value of the target review strategy.

To illustrate the target review strategies in the successive model, consider the example from earlier in this section. Suppose the review ratio on day 0 is 0.5 and the target review ratio is 0.6. Then the review ratios on days 1 to 5 equal 0.5, 0.5, 0.5, 0.6, and 0.6, respectively. All performances are then separately solved and revenue is optimised to target review ratio 0.6.

In the remainder of this section, a method is developed that optimises revenue to a target review ratio.

3.3. One Performance Dynamics

The one performance problem can be modelled as a finite-horizon continuous-time multi-objective Markov decision process over T time units. Define the state space by

$$X := \{0, \ldots, C_{\max}\},$$

where $x \in X$ represents the number of reservations. The *action space* $\mathcal{A}$ at each time step is the set of all possible offer sets $S \subset N$.

To solve the problem, time is discretised and divided into T time periods, where the length of the intervals is such that the probability that more than one event occurs is very small (following [17]). To formulate the *transition probabilities*, it is assumed (following [17]) that only one event occurs per time period, where an event is either an arrival, a cancellation, or neither arrival nor cancellation. Denote with λ the probability that a customer arrives in a time period; and γx the probability that a product is cancelled in state x. The probability that no purchase occurs in a time period equals the sum of the probability that neither an arrival and nor a cancellation occurs, and the probability that an arrival occurs but the arriving customer makes no purchase. This is equal to

$$(1-\lambda-\gamma x)+\lambda P_0(S)=1-\lambda\sum_{j\in S}P_j(S)-\gamma x.$$

In each time period the decision needs to be made which set S to offer. Note that time has to be scaled such that $\lambda+\gamma C\leq 1$ for the probabilities to be well defined. The transition probabilities from a given state $0<x<C_{max}$, are thus:

$$\begin{aligned}
P(x'=x+1|x,S) &= \lambda\sum_{j\in S}P_j(S),\\
P(x'=x-1|x,S) &= \gamma x,\\
P(x'=x|x,S) &= 1-\lambda\sum_{j\in S}P_j(S)-\gamma x,
\end{aligned}$$

and 0 otherwise. When $x=C_{max}$ it is not allowed to offer any fare products, and the only possible transitions are due to cancellations.

Now define the *reward function* $\mathbf{R}\colon X\times\mathcal{A}\to\mathbb{R}^3$ as follows. Let $R^1(x,S)$ be the expected immediate revenue reward; $R^2(x,S)$ the expected good review reward; and $R^3(x,S)$ the expected bad review reward, in a state x for an action S. Define $r_{1j}=r_j$, $r_{2j}=q_j^p$, and $r_{3j}=-q_j^n$ for all $j\in N$. Define $c_{1j}(t)=c_j(t)$, $c_{2j}(t)=q_j^p$, and $c_{3j}(t)=-q_j^n$ for all $j\in N$ and for all t. The expected costs that follow from cancellations can be added to the immediate rewards when a product is purchased, since neither the booking system nor the manager has control over the cancellations of current reservations (see [17] for a detailed description of this derivation). By doing so, it is not necessary to account for the costs of cancellations of a product purchased at a time step later in the decision process, allowing the state x to be Markov.

The expected cancellation costs for objective $o\in\{1,2,3\}$ that follow from selling product j in time step t is equal to

$$\Delta H_{oj}(t)=\begin{cases}\gamma c_{oj}(t)+(1-\gamma)\Delta H_{oj}(t-1) & \text{if } t>1,\\ 0 & \text{if } t=1,\end{cases}$$

The expected reward that includes the expected costs of cancellations is therefore given by $r_{oj}-\Delta H_{oj}(t)$, for all $j\in N$, for $o=1,2,3$, and for all t. The immediate reward for a cancellation[4], as well as nothing

[4]The immediate reward for cancellations is 0, because the expected costs are already accounted for in the purchase

happening is $\mathbf{0}$.

3.4. Policies and Value Vectors

A manager can interact with the model by executing a *policy* with a corresponding *value*. In the multi-objective rather than single-objective setting, stochastic policies can Pareto dominate deterministic polices [21, 23]. Therefore, define the policy space Π, as all possible mappings of states, time steps, and actions (i.e., fare-products) to a probability of taking that action: $\pi\colon X \times \{0,\dots,T\} \times \mathcal{A} \to [0,1]$. For each state and time step, the probabilities for each action should be positive and should sum to 1.

A given policy, $\pi(x,t,S)$, induces a probability distributions over execution trajectories $(x_T, S_T, \mathbf{r}_T, \ldots$ $x_1, S_1, \mathbf{r}_1, x_0)$, where $\mathbf{r}_t$ is the reward vector corresponding to time step t. Each such a trajectory has an associated vector-valued *return*, i.e., the sum of the reward vectors in the trajectory. The value of a policy is its expected return. Because the returns are additive [14], the value of a policy π for a given time step and state, $\mathbf{V}^{\pi}_{t}(x)$ can be expressed recursively with a Bellman equation [1] in vector form, or, as separate Bellman equations per objective:

$$\begin{aligned} V^{\pi}_{o,t}(x) = \sum_{S\subset N} \pi(x,t,S) \Bigg\{ & \lambda \sum_{j\in S} P_j(S) \left[r_{oj} - \Delta H_{oj}(t) + V^{\pi}_{o,t-1}(x+1) \right] \\ & + \gamma x V^{\pi}_{o,t-1}(x-1) \\ & + \left(1 - \lambda \sum_{j\in S} P_j(S) - \gamma x \right) V^{\pi}_{o,t-1}(x) \Bigg\}. \end{aligned} \tag{1}$$

A special kind of policy is the deterministic policy, i.e., policies in which one action per state and time step has probability 1. A deterministic policy $\pi(x,t)$ may be seen as a mapping: $\pi\colon X \times \{0,\dots,T\} \to \mathcal{A}$. In this case, the marginalisation over actions drops out of the Bellman equation:

$$\begin{aligned} V^{\pi}_{o,t}(x) =& \lambda \sum_{j\in S} P_j(\pi(x,t)) \left[r_{oj} - \Delta H_{oj}(t) + V^{\pi}_{o,t-1}(x+1) \right] \\ & + \gamma x V^{\pi}_{o,t-1}(x-1) \\ & + \left(1 - \lambda \sum_{j\in S} P_j(\pi(x,t)) - \gamma x \right) V^{\pi}_{o,t-1}(x). \end{aligned} \tag{2}$$

event rewards [17].

There is typically no single policy that is best with respect to all objectives, because there typically are different policies that yield different trade-offs between the objectives. In the multi-objective decision making literature, a certain trade-off is typically preferred on the basis of the partially unknown preferences of a user, i.e., a human decision maker [12]. These preferences can be expressed in terms of a *scalarisation function*, or *utility function*, f, which collapses the value vector of a policy, i.e., the vector of values of a policy in each objective, to a scalar:

$$V_{\mathbf{w}}^{\pi} = f(\mathbf{V}^{\pi}, \mathbf{w}), \tag{3}$$

where $\mathbf{V}^{\pi}$ is the value vector of a policy π, and $\mathbf{w}$ is a vector parametrising f. Typically, only limited information is available about the function f. Therefore, the solution to a multi-objective decision problem is a set of value vectors and associated policies that *cover* all possible f and $\mathbf{w}$. The human decision maker can choose her preferred policy from this set.

In our problem, there is no human decision maker. Instead, the function f represents the long-term revenue, as a result of optimally balancing the review scores and the immediate revenue. To determine the optimal long-term revenue exactly however, the intractable problem discussed at the beginning of this section has to be solved. Gladly, it is not necessary to do that at this stage. Instead, assume that f is unknown but with some known constraints. This results in a tractable model that describes one-day-arrival dynamics.

In the problem at hand, i.e., the one-day-arrival dynamics multi-objective Markov decision process (MOMDP) model, the exact shape of f and $\mathbf{w}$ are unknown. Therefore, the aim is to find a set of all possibly optimal solutions for the MOMDP, i.e., a set that *covers* all possible f and $\mathbf{w}$ that fit the known constraints about f and $\mathbf{w}$ [14]:

Definition 3.1. *For a multi-objective Markov decision process m and a family of scalarisation functions $\mathcal{F}$, a set of policies CS is a* coverage set *if for every $\mathbf{w}$, it contains a policy that has the maximal scalarised value, i.e., if:*

$$(\forall f{\in}\mathcal{F}, \mathbf{w})(\exists \pi \in CS)\left(\forall(\pi' \in \Pi^m)\ V_{\mathbf{w}}^{\pi} \geq V_{\mathbf{w}}^{\pi'}\right) \wedge (\forall \pi \in CS)(\exists f{\in}\mathcal{F}, \mathbf{w})\left(\forall(\pi' \in \Pi^m)\ V_{\mathbf{w}}^{\pi} \geq V_{\mathbf{w}}^{\pi'}\right), \tag{4}$$

where Π^m is the set of all valid policies for m.

Typically, the aim is to make the coverage set as small as possible; the smaller the CS, the easier for the user to compare the alternatives. However, even if only undominated policies are considered, i.e., policies that are optimal for at least one f and $\mathbf{w}$, the CS may not be unique; multiple policies may cover the same subset of possible f and $\mathbf{w}$. In the next section, the CS is specified further, by imposing constraints on f that follow from the available information about how the value vectors attainable in the MOMDP model affect the long-term revenue. After that, the coverage set for the one-day-arrival dynamics MOMDP is used to determine approximately optimal strategies for the full problem.

4. Computing Coverage Sets for Single-performance Planning

In this section, coverage sets for one-day-arrival dynamics MOMDP planning are discussed. Firstly, the appropriate CS for is derived from the available information about how immediate revenue and reviews affect the long-term revenue. Secondly, an algorithm that computes this CS is discussed. Finally, a way to employ the CS to improve long-term revenue is proposed.

4.1. Coverage Sets

Recall that the long-term revenue can be expressed in terms of a scalarisation function $f(\mathbf{V}^\pi, \mathbf{w})$. Which undominated policies need to be contained in the coverage set depends on what is known about f and $\mathbf{w}$. Sometimes it is known beforehand that f has a particular shape. For example, f might be *linear* [24].

Definition 4.1. *A linear scalarisation function is the inner product of a weight vector* $\mathbf{w}$ *and a value vector* $\mathbf{V}^\pi$

$$V^\pi_{\mathbf{w}} = \mathbf{w} \cdot \mathbf{V}^\pi. \tag{5}$$

Each element of $\mathbf{w}$ *is greater than or equal to* 0*, and specifies how much one unit of value for the corresponding objective contributes to the scalarised value.*

Linear scalarisation functions are both common and intuitive. The most common situation in which linear scalarisation applies is when the value vectors can be translated into monetary value. For example, consider a task in which objective corresponds to quantities of various resources that need to be bought or sold on a market. For revenue management this might therefore be the most intuitive way of scalarising. However, linearity is a strong assumption, that cannot be made in this context.

In fact, only the following information about f can safely be assumed:

- immediate revenue always contributes positively to the long-term revenue,
- positive reviews contribute positively to the long-term revenue, and
- negative reviews always contribute negatively to the long-term revenue.

When the negative review objective is redefined as -1 times the number of negative reviews, an f is obtained that is *monotonically increasing* in all objectives.

Definition 4.2. *A scalarisation function f is* monotonically increasing *if:*

$$\forall o,\ V_o^{\pi} \geq V_o^{\pi'} \Rightarrow \forall \mathbf{w},\ f(\mathbf{V}^{\pi}, \mathbf{w}) \geq f(\mathbf{V}^{\pi'}, \mathbf{w}), \tag{6}$$

where V_o^{π} denotes the value of policy π in the o-th objective.

The assumption that is f is *monotonically increasing* guarantees that if more of one objective is obtained while not losing anything in another objective, the utility cannot go down. Note that linear scalarisation functions (with non-zero positive weights) are included in the family of monotonically increasing functions. Monotonicity is therefore a less strict assumption than linearity.

When an optimal solution is needed with respect to all linear f, a coverage set is needed that contains an optimal solution for every possible weight vector $\mathbf{w}$ in Equation 5. The coverage with respect to linear f is called a *convex coverage set (CCS).*

Definition 4.3. *A set of policies CCS is a* convex coverage set *if:*

$$\forall \mathbf{w}\ \exists \pi \in CCS \left(\forall (\pi' \in \Pi^m)\ \mathbf{w} \cdot \mathbf{V}^{\pi} \geq \mathbf{w} \cdot \mathbf{V}^{\pi'} \right) \wedge \forall \pi \in CCS\ \exists \mathbf{w} \left(\mathbf{w} \cdot \mathbf{V}^{\pi} \geq \mathbf{w} \cdot \mathbf{V}^{\pi'} \right). \tag{7}$$

Without loss of generality, assume that $\mathbf{w}$ adheres to the simplex constraints, i.e., all elements of $\mathbf{w}$ are positive and sum to one.

The coverage set for the family of monotonically increasing functions is called a *Pareto coverage set (PCS).*

Definition 4.4. *A set of policies PCS is a* Pareto coverage set *if:*

$$\forall \pi' \in \Pi^m \; \exists (\pi \in PCS) \left(\mathbf{V}^{\pi} \succeq_P \mathbf{V}^{\pi'} \right) \tag{8}$$

where $\succeq_P$ denotes weak Pareto-dominance, i.e., $\mathbf{V}^{\pi} \succeq_P \mathbf{V}^{\pi'} \equiv \forall o \;\; V_o^{\pi} \geq V_o^{\pi'}$.

I.e., for every possible policy, there is a policy in the PCS, with at least equal value in all objectives.

Because monotonicity is a less strict assumption than linearity, the PCSs are typically much larger than the CCSs. Furthermore, because of the possibly non-linearity of the scalarisation function, policies that constitute a PCS are much harder to obtain than those that constitute a CCS. Gladly however, it is not necessary to represent PCSs explicitly if it can be assumed that stochastic policies are allowed. This is due to a theorem by Vamplew et al. [21] which states that for any MOMDP, a PCS of stochastic policies can be constructed from a given CCS of deterministic policies, by taking so-called *mixture policies* from policies in this CCS. A mixture policy is constructed by taking a subset of N policies from the CCS and assigning each a probability of being executed.

The difference between the CCS and the PCS with or without stochastic policies is illustrated in Figure 4, where all the possible values of deterministic policies for a 2-objective MOMDP are denoted as points. The axes represent the different objectives. Note that the grey points are neither in any PCS or CCS, as they are dominated by one of the other points (i.e., there is another point with a higher value in all objectives. The red points represent a possible CCS of deterministic policies, while the blue point represents a point that may be, but is not necessarily, in the CCS. Note that a CCS of deterministic policies is also a CCS for stochastic policies as there is always a deterministic policy that is optimal for any $\mathbf{w}$ [6, 14]. The black point would be in a PCS if stochastic policies are not allowed as it is not dominated by another point. However, when stochastic policies (including mixture policies) are allowed, the PCS is represented by all the values on the red lines, on which there are policy values that dominate the black point. To summarise, all the Pareto-optimal policy values of stochastic policies can be constructed by mixing deterministic policies from a CCS of deterministic policies. Therefore, this

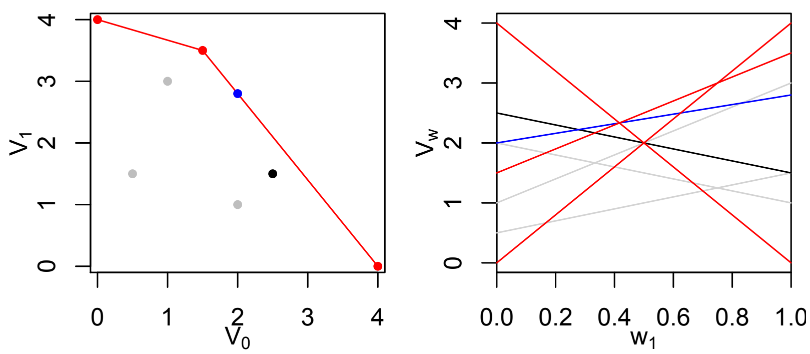

Figure 4: *(Left)* A stochastic PCS can be constructed from a deterministic CCS. *(Right)* The *scalarised value* as a function of the linear scalarisation weight.

article focuses on finding methods for finding a CCS, and construct policies with values on the PCS of stochastic policies when necessary. Thus, define solving an MOMDP as *finding a convex coverage set.*

4.2. Optimistic Linear Support

In order to compute the CCS for the single arrival model, the *Optimistic Linear Support (OLS)* algorithm [15] is used. OLS is a general framework for solving multi-objective decision problems (including MOMDPs [13]). It takes a single-objective solver — such as dynamic programming — as a subroutine, and produces a CCS within a finite number of calls to this subroutine. In this section, the framework is described, and the specific implementation of OLS for multi-objective revenue management.

Because OLS computes the CCS, it can make use of linear scalarisation. Under this assumption, define the *scalarised value function* $V^*_{CCS}(\mathbf{w})$ that provides the maximal scalarised value given a linear scalarisation weight $\mathbf{w}$:

$$V^*_{CCS}(\mathbf{w}) = \max_{\mathbf{V}^\pi \in CCS} \mathbf{w} \cdot \mathbf{V}^\pi.$$

$V^*_{CCS}(\mathbf{w})$ is a piecewise linear and convex (PWLC) function, because each value function defines a (hyper)plane over the weight simplex, as illustrated in Figure 4 (right), and $V^*_{CCS}(\mathbf{w})$ maximises over these hyperplanes. That is, it consists of the convex upper surface of the lines in Figure 4 (right).

OLS builds up the CCS incrementally by solving a series of linearly scalarised instances of the MOMDP, for different $\mathbf{w}$. The optimal policy π to an instance scalarised with $\mathbf{w}$ maximizes $V^\pi_{\mathbf{w}} = \mathbf{w} \cdot \mathbf{V}^\pi$. When this π is identified, $\mathbf{V}^\pi$ is added to a partial CCS X, which converges to a CCS.

In order to select good $\mathbf{w}$'s for scalarisation, OLS exploits the observation that $V^*_X(\mathbf{w}) = \max_{\mathbf{V}^\pi \in X} \mathbf{w} \cdot \mathbf{V}^\pi$ is PWLC over the weight simplex. In particular, OLS selects only so-called *corner weights* that lie at the intersections of line segments of the PWLC function $V^*_X(\mathbf{w})$ that correspond to the value vectors found

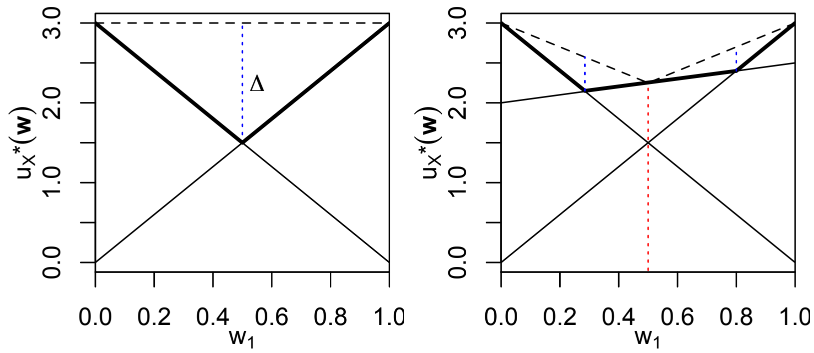

Figure 5: *(Left)* The scalarised value as a function of weights $V_X^*(\mathbf{w})$ (bold segments) for $X = \{(0,3),(3,0)\}$. There is one corner weight: $(0.5, 0.5)$. *(Right)* Adding a new value vector, $(2.0, 2.5)$, to X, thereby improving $V_X^*(\mathbf{w})$.

so far. For example, in Figure 5 on the left, $V_X^*(\mathbf{w})$ is indicated with bold line segments. There are two value vectors in X, and there is one corner weight. The maximal potential error reduction that can be made by identifying a new value vector $\mathbf{u}(\mathbf{a})$ is at a corner weight [3]. The potential error reduction is denoted with dashed blue vertical lines, and is at present Δ. OLS scalarises the MOMDP at this corner weight and solves it using a single-objective solver, obtaining the optimal policy for that weight. The multi-objective value of this policy, $\mathbf{V}^\pi$, improves over $V_X^*(\mathbf{w})$ at that corner weight, as shown on the right-hand side of Figure 5 indicated by the red dashed line on the right. By continuing to find new corner weights and solving scalarised MOMDPs for corresponding to these corner weights, OLS is guaranteed to produce an exact CCS after solving a finite number of single-objective problems. That is, when no improvements are found at any remaining corner weights, the possible error reduction becomes 0, and a CCS is found. When there are multiple corner weights, OLS first solves the scalarised problem with the highest possible error reduction Δ.

After solving a scalarised MOMDP, for a given $\mathbf{w}$, a policy $\mathbf{w}$ is obtained. If a standard implementation of dynamic programming would be used, the single-objective policy value $V_\mathbf{w}^\pi$ would also be obtained. Because OLS requires $\mathbf{V}^\pi$ however, this would mean that policy evaluation has to be used to obtain this multi-objective value. Therefore, in this implementation, an improvement is made over standard DP which is called *scalarised dynamic programming (SDP)*. SDP keeps track of the multi-objective value vectors while maximising the value for $\mathbf{w}$, thereby preventing having to perform separate policy evaluation steps.

4.3. Equilibrium Strategy

This article proposes to find target review strategies that maximise revenue in the long run. For a given CCS with input review ratio ρ, a target review strategy $\pi(\rho)$ that achieves ρ can be found as follows. Each deterministic stationairy policy $\pi \in$ CCS has a corresponding value vector $\mathbf{V}^\pi$. Hence each policy can be represented in terms of a value vector. This leads to a three-dimensional value space CCS$'$ with revenue, positive reviews, and negative reviews as axes:

$$\mathrm{CCS}' = \{\mathbf{V}^{\pi} \,|\pi \in \mathrm{CCS}\}. \tag{9}$$

Consider Figure 6 below, where a sketch of CCS′ is given by the convex non-linear surface. This surface represents all potential optimal value vectors, covering all review ratios. However, not all value vectors need to be considered: the space CCS′ can be reduced to the value vectors $\mathbf{V}^{\pi} \in \mathrm{CCS}'$ for which the review ratio equals ρ. In order to do so, consider the hyperplane $\mathcal{H}$ for which the review ratio equal ρ:

$$\mathcal{H} = \left\{ x \in \mathbb{R}^3 \,\middle|\, \frac{x_2}{x_2 - x_3} = \rho \right\} = \left\{ x \in \mathbb{R}^3 \middle| \, x_2(\rho - 1) - \rho x_3 = 0 \right\}.$$

A sketch of the hyperplane is given in Figure 6. The reduced value vector space, consisting of value vectors in CCS′ for which the review ratio is equal to ρ, is then given by the intersection $\mathrm{CCS}' \cap \mathcal{H}$. In Figure 6 the intersection CCS′ and $\mathcal{H}$ is empasised by the black thick line.

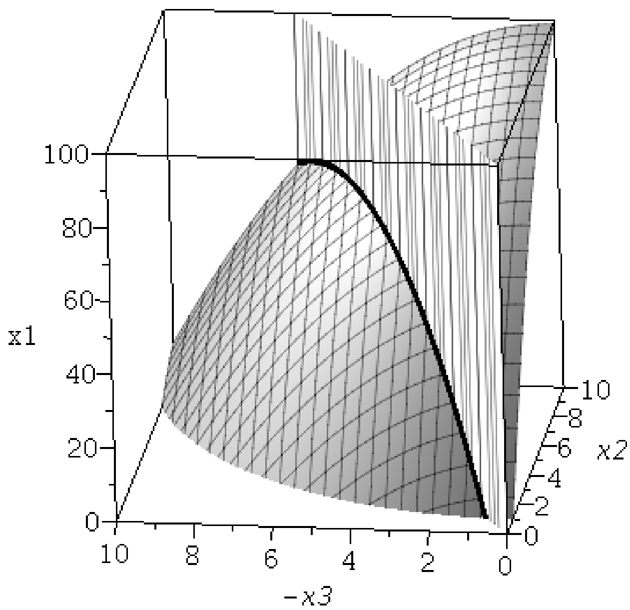

Figure 6: Visualisation of the set CCS′ of potentially optimal value vectors (convex non-linear surface), the hyperplane $\mathcal{H}$ of value vectors for which the review ratio is equal to ρ (hyperplane cutting CCS′), and their intersection $\mathrm{CCS}' \cap \mathcal{H}$, the set of all potentially optimal value vectors for which the review ratio equals ρ (thick black line).

Hence the value vector $\mathbf{V}^{\pi} \in \mathrm{CCS}' \cap \mathcal{H}$ with optimal revenue gives the solution to the problem at hand:

$$\pi(\rho) = \underset{\pi \in \mathrm{CCS}}{\arg\max} \left\{ V_1^{\pi} \,\middle|\mathbf{V}^{\pi} \in \mathrm{CCS}' \cap \mathcal{H} \right\}.$$

A closer look at the solution space provides insight in the evaluation of the optimal solution $\pi(\rho)$. First note that CCS, and therefore CCS$'$, is a convex set consisting of a finite number of faces. Therefore, CCS$' \cap \mathcal{H}$ is a line in $\mathbb{R}^3$ consisting of a finite number of line segments, corresponding to faces of CCS$'$ that intersect with $\mathcal{H}$. Hence $\pi(\rho)$ is one of the corners of those line segments.

Now that a target review strategy $\pi(\rho)$ can be found for every ρ, the goal is to find the strategy π^*, with corresponding target review ratio ρ^*, that optimises revenue:

$$\pi(\rho^*) = \arg\max_{\pi(\rho)} V_1^{\pi(\rho)}.$$

Two computational challenges arise in these evaluations:

- Not all target review ratios are attainable. The CCS with input ρ can attain the review ratios $[\underline{\rho}, \overline{\rho}]$. If $\rho \notin [\underline{\rho}, \overline{\rho}]$, then no target review strategy exists and ρ is not a feasible target review ratio. Only feasible target review ratios need to be considered. However, it is not clear beforehand which values of ρ are feasible.

- Optimising over all target review ratios is challenging. The set of feasible target review ratios is continuous, and in order to find the optimal revenue corresponding to a certain target review ratio ρ the whole procedure described in this section has to be followed. Under these deliberations the authors propose two approaches to deal with the continuous variable ρ. The first approach descretises the set of attainable values of ρ, denoted by $\mathcal{P} \subset \mathbb{R}$, $|\mathcal{P}| < \infty$. Then, an optimal policy $\pi(\rho^*)$ over $\mathcal{P}$ is used:

  $$\rho^* = \arg\max_{\rho \in \mathcal{P}} V_1^{\pi(\rho)}. \tag{10}$$

  The second approach is iterative and of stochastic nature. Each iteration an arbitrary ρ is selected from the set of feasible values for ρ, according to some distribution. If $V_1^{\pi^*(\rho)} > V_1^{\pi^*(\rho')}$, with ρ' the best target review ratio found so far, then update ρ' with ρ. Continue until some stopping criterion is hit.

Remark: Traditional RM optimisation models provide intuitive and interpretable strategies, such as booking limits or protection levels [20]. However, such a nice structure is absent when the model becomes more complicated, such as when cancellations and customer choice behaviour are combined [17]. The complexity of our model goes even beyond cancellations and customer behaviour, so unfortunately a simple booking limit or protection level structure is impossible in our case.

5. Numerical Examples

In this section numerical results are provided of an implementation of the multi-objective revenue management model. The two main goals of these examples are 1) to show how to interpret and use the convex coverage set; 2) to numerically validate target review strategies. Examples of realistic size are used. All examples use the set-up of the running example introduced in Section 2.

5.1. One Instance: Policy Analysis

This example examines the results for scenario 1 from the example of Section 2, where the current review ratio is $\rho = 0.6$. The resulting convex coverage set is presented in Table 2. Besides the total expected revenue, the positive reviews, and the negative reviews, also the resulting review ratio ρ is given. The solutions are ordered by revenue. The value space is three dimensional (revenue, positive reviews, and negative reviews). To provide a graphical representation of the value space, consider the three faces presented in Figures 7, 8, and 9. In each figure the convex coverage set that follows from revenue and review ratio is also given. Moreover, Figure 10 provides a visual representation of review ratio versus revenue.

Two observations of interest can be made from this table and accompanied figures. First, by sacrificing revenue the review ratio can be increased, as is expected. Optimising revenue results in total expected revenue of 27098.09 and review ratio of 0.57, while optimising the review ratio yields total expected revenue of only 7493.76, a 72% decline, and review ratio of 0.9, a 56% increase. Sacrificing revenue does not necessarily lead to a higher review ratio though. For example, the optimal solution corresponding to the sixth row in Table 2 leads to an expected revenue of 15137.33 and review ratio 0.84. The fifth row shows that by sacrificing revenue to 14252.8 not only the number of positive reviews increase, but also the number of negative reviews increase. In this case it leads to a decrease in review ratio to 0.81. Therefore, the procedure in Section 3 needs to be used to find a target review strategy. In this case, The target review strategy of $\rho^* = 0.6$ has optimal revenue of 27027.18, a 3.18% increase with respect to solely optimising revenue (at $\rho = 0.57$).

A second observation is that the positive reviews tend to increase as revenue decreases, and negative reviews tend to decrease as revenue decreases. This is in concordance with the set-up of the scenarios, where higher prices lead to less positive and more negative reviews. It can also be seen in Figures 7 and 8. However, there is no strict increase or decrease. This can be explained by the trade-off between positive and negative reviews with respect to revenue. There is however a difference in behaviour between positive and negative reviews. At first positive reviews tend to increase as revenue decreases. Policies are selected that give slightly less revenue, but provide more positive reviews and less negative reviews. However, at

Revenue	Positive reviews	Negative reviews	Rating ρ
0	0	0	-
7493.76	33.01	3.75	0.90
12513.99	42.88	9.28	0.82
12595.63	41.42	6.30	0.87
14252.80	42.48	9.77	0.81
15137.33	41.09	7.57	0.84
16021.48	41.67	10.38	0.80
17983.01	40.40	11.16	0.78
18693.24	39.78	10.61	0.79
19985.47	38.61	9.99	0.79
21719.66	36.79	11.74	0.76
23528.79	33.55	12.81	0.72
24686.86	26.26	12.34	0.68
25194.50	29.93	13.87	0.68
26943.83	25.33	15.07	0.63
27098.09	21.11	15.62	0.57

Table 2: Convex coverage set for $\rho = 0.6$.

some point revenue can only be increased more if at times no products are offered, to the point that no products are offered any more. This leads to the zero-solution: if no products are offered, no revenue is earned and no reviews are given. This in fact is an optimal policy when only negative reviews are the objective. A side-effect is that at some point, when revenue decreases, also positive reviews decrease, because of lack of offer. For example, consider the solution with 12513.99 revenue, 42.88 positive reviews, and 9.28 negative reviews; compared to revenue of 7493.76, 33.01 positive reviews, and 3.75 negative reviews. Note that the reviews score does increase from 0.82 to 0.90.

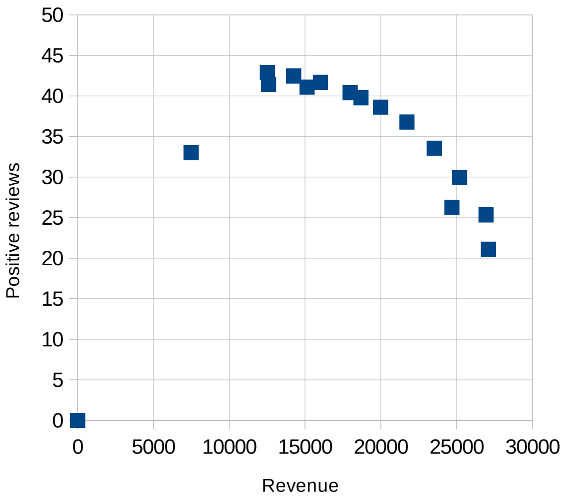

Figure 7: Convex coverage set plot of revenue and positive reviews.

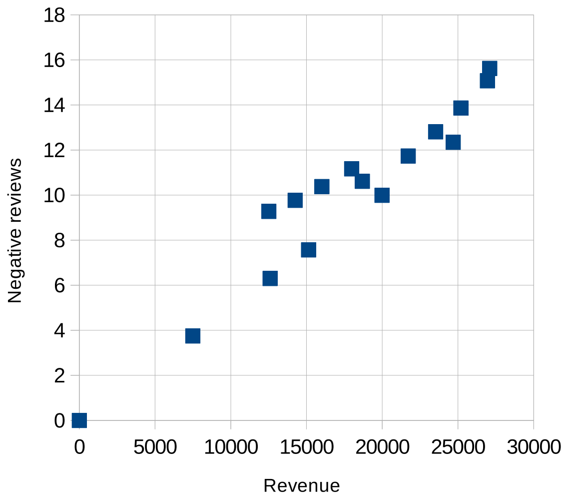

Figure 8: Convex coverage set plot of revenue and negative reviews.

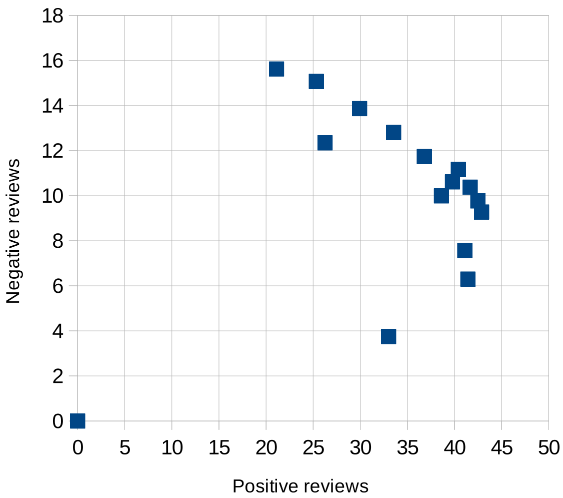

Figure 9: Convex coverage set plot of positive and negative reviews.

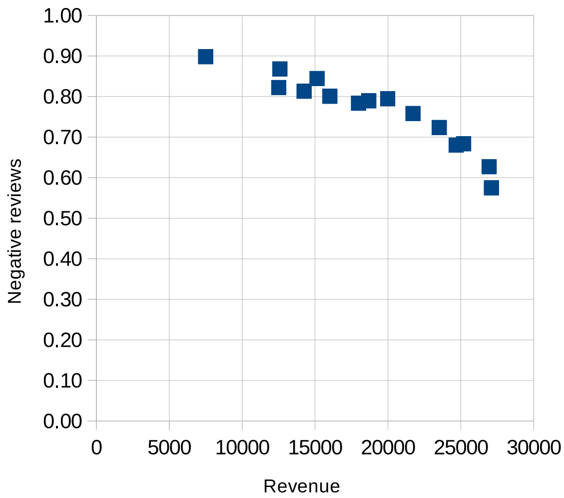

Figure 10: Plot of revenue and review ratio resulting from CCS.

An important application of the revenue-review trade-off model is to make decisions now that positively influence future purchase behaviour and revenues. By sacrificing revenue now in order to get better reviews, future revenues can be increased substantially. One way to do this is to consider the target review strategies described in Section 3 (other approaches that are both tractable and indicate how long-term revenue can be increased are not known to the authors). Each of the four scenarios of the example of Section 2 is considered. The optimal revenue that results from the target review ratio can be calculated using stochastic mixture policies on the convex coverage set of revenue and reviews, as is shown in Section 3. For example, the target review strategy for scenario 1 with $\rho = 0.6$ yields an expected revenue of 27027.18. Just optimising revenue, on the other hand, leads to a target review policy that yields an expected revenue of 26194.23. By observing the policy space in Table 2, an initial revenue *loss* of 0.26% is incurred (from 27098.09 if solely revenue is optimised). However, on the long term this leads to a revenue *increase* of 3.18%.

The question now is what target review ratio will lead to maximal revenue in the long term. Therefore the following experiment is conducted. For all four scenarios the total expected revenue corresponding to the target review policy is evaluated for a range of target review ratios. For visual convenience, losses for some higher target reviews are omitted. The results are presented in Figure 11. Observe that the curves are not smooth, nor indeed is there a reason why they should be.

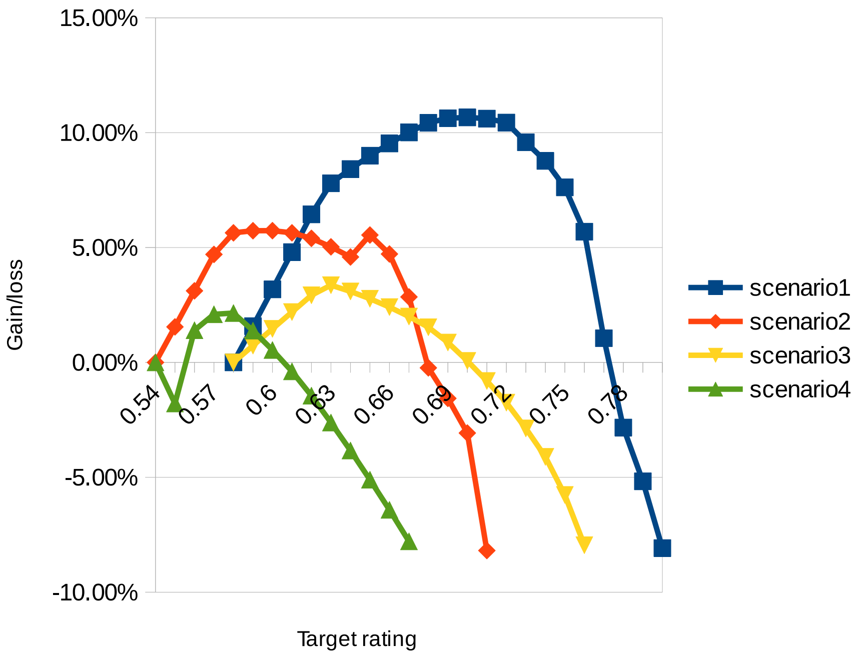

Figure 11: Revenues corresponding to target review policies.

The four scenarios give different results, though the prevailing observation is that there is a lot to gain by optimising to target review ratio instead of solely optimising revenue. Scenario 1, where both the effect of review ratio on purchases and of purchases on review ratio are large, shows a structural increase in revenue of 11% when an optimal target review ratio is selected. When demand effect is small, but the effect of review probabilities high, in scenario 2, then optimising to target review ratio leads to a 6% structural increase in revenue. The other scenarios 3 and 4, with the effect of review probabilities low, show a decent increase in revenue of about 2-3%.

Research in [25] found that a 10% increase in rating increases online bookings by more than 5%. In terms of revenue, our example gives similar results. Scenario 1 induces a revenue increase of 10.66% with an increase in review ratio of 20.69%. The increase in revenue is higher that in the results by [25], but so is the increase in review ratio. In scenario 2 a similar observation can be made: 5.73% increase in revenue and 11.11% increase in review ratio. Scenario 3 and 4, however, show slightly poorer results. Scenario 3 yields an increase of 3.36% in revenue with 8.62% increase in review ratio and scenario 4 an increase of 2.15% in revenue and 7.41% in review ratio. This can be explained by the low impact of review ratio on demand in these scenarios.

5.3.

6. Concluding Remarks

In this article a novel revenue management model is introduced that captures the trade-off between revenue and reviews. Revenue management strategies influence the perception of customers, which results in a changing review ratio. On the other hand, review ratio influences buying behaviour. The formulated model captures the long-term effect of optimising revenue according to a target review ratio.

An innovative solution method to approach a problem of such complexity is introduced to optimise revenue in the long run. The methodology builds on recent developments in multi-objective Markov decision process theory, and contributes to this body of literature. Because the policy space is restricted to policies where revenue is optimised such that the target review ratio remains constant, the full problem can be reduced to a series of multi-objective Markov decision problems.

The numerical studies show how to interpret the solution space, the convex coverage set, or a single multi-objective MDP in the series. Moreover, results of the target review strategies of the successive model show that revenue improvements of up to 11% are achievable if reviews are taken into account in the optimisation process, instead of the sole objective of revenue. All results, featuring different scenarios, suggest revenue increases of at least 2%. In practical terms for theatre, this leads to a significant increase in revenue that can reach into the millions annually.

The results of this article have several implications that lead to topics for further research. First, the model can be used to identify the effect of improving facilities of the theatre to revenue and review ratios. Second, the model can be extended to a network setting for applications with such structure, like hotels. However, this is not straightforward, and it increases the dimensionality of the problem to the extend that is intractable.

[1] R.E. Bellman. A Markov decision process. *Journal of Mathematical Mech.*, 6:679–684, 1957.

[2] M. Ben-Akiva and S.R. Lerman. *Discrete Choice Analysis.* MIT Press, Cambridge, MA, 1985.

[3] Hsien-Te Cheng. *Algorithms for partially observable Markov decision processes.* PhD thesis, University of British Columbia, Vancouver, 1988.

[4] C. Dellarocas, N. Awad, and X. Zhang. Exploring the value of online reviews to organizations: Implications for revenue forecasting and planning. *ICIS 2004 Proceedings*, 30, 2004.

[5] Katja Hofmann. *Fast and reliable online learning to rank for information retrieval.* PhD thesis, University of Amsterdam, 2013.

[6] R.A. Howard. *Dynamic programming and Markov decision processes.* MIT Press, 1960.

[7] Thorsten Joachims, Laura Granka, Bing Pan, Helene Hembrooke, Filip Radlinski, and Geri Gay. Evaluating the accuracy of implicit feedback from clicks and query reformulations in web search. *ACM Transactions on Information Systems (TOIS)*, 25(2):7, 2007.

[8] Conrad J Lautenbacher and Shaler Stidham Jr. The underlying markov decision process in the single-leg airline yield-management problem. *Transportation Science*, 33(2):136–146, 1999.

[9] Y. Pan and J. Q. Zhang. Born unequal: a study of the helpfulness of user-generated product reviews. *Journal of Retailing*, 87(4):598–612, 2011.

[10] C. Park and T. M. Lee. Information direction, website reputation and ewom effect: A moderating role of product type. *Journal of Business Research*, 62(1):61–67, 2009.

[11] P. A. Pavlou and A. Dimoka. The nature and role of feedback text comments in online marketplaces: Implications for trust building, price premiums, and seller differentiation. *Information Systems Research*, 17(4):392–414, 2006.

[12] Diederik M Roijers and Shimon Whiteson. Multi-objective decision making. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 11(1):1–129, 2017.

[13] D.M. Roijers, J. Scharpff, M.T.J. Spaan, Frans A. Oliehoek, Mathijs de Weerdt, and Shimon Whiteson. Bounded approximations for linear multi-objective planning under uncertainty. In *ICAPS 2014:*

Proceedings of the Twenty-Fourth International Conference on Automated Planning and Scheduling, pages 262–270, 2014.

[14] D.M. Roijers, P. Vamplew, S. Whiteson, and R. Dazeley. A survey of multi-objective sequential decision-making. *Journal of Artificial Intelligence Research*, 48:67–113, 2013.

[15] D.M. Roijers, S. Whiteson, and F.A. Oliehoek. Computing convex coverage sets for faster multi-objective coordination. *Journal of Artificial Intelligence Research*, 52:399–443, 2015.

[16] Z.M. Shen and X. Su. Customer behavior modeling in revenue management and auctions: A review and new research opportunities. *Production and Operations Management*, 16(6):713–728, 2007.

[17] D.D. Sierag, G.M. Koole, R.D. van der Mei, J.I. van der Rest, and B. Zwart. Revenue management under customer choice behaviour with cancellations and overbooking. *European Journal of Operational Research*, 246(1):170–185, 2015.

[18] B.A. Sparks and V. Browning. The impact of online reviews on hotel booking intentions and perception of trust. *Tourism Management*, 32(6):1310–1323, 2011.

[19] K. Talluri and G. van Ryzin. Revenue management under a general choice model of consumer behavior. *Management Science*, 50(1):15–33, 2004.

[20] K. T. Talluri and G. J. Van Ryzin. *The theory and practice of revenue management*, volume 68. Springer Science & Business Media, 2006.

[21] P. Vamplew, R. Dazeley, E. Barker, and A. Kelarev. Constructing stochastic mixture policies for episodic multiobjective reinforcement learning tasks. In *AI'09: The 22nd Australasian Conference on Artificial Intelligence*, pages 340–349, 2009.

[22] I. E. Vermeulen and D. Seegers. Tried and tested: The impact of online hotel reviews on consumer consideration. *Tourism Management*, 30(1):123–127, 2009.

[23] K. Wakuta. A note on the structure of value spaces in vector-valued Markov decision processes. *Mathematical Methods of Operations Research*, 49(1):77–85, 1999.

[24] C.C. White and K.M. Kim. Solution procedures for solving vector criterion Markov decision processes. *Large Scale Systems*, 1:129–140, 1980.

[25] Q. Ye, R. Law, B. Gu, and W. Chen. The influence of user-generated content on traveler behavior: An empirical investigation on the effects of e-word-of-mouth to hotel online bookings. *Computers in Human Behavior*, 27(2):634–639, 2011.

[26] K. H. Yoo and U. Gretzel. Influence of personality on travel-related consumer-generated media creation. *Computers in Human Behavior*, 27(2):609–621, 2011.

[27] L. Zhou, S. Ye, P. L. Pearce, and M. Y. Wu. Refreshing hotel satisfaction studies by reconfiguring customer review data. *International Journal of Hospitality Management*, 38:1:10, 2014.