

Risk Minimizing Strategies for Revenue Management Problems with Target Values

Matthias Koenig

*Department of Management Science, Lancaster University Management School, United Kingdom.
m.koenig@lancaster.ac.uk*

Joern Meissner

Kuehne Logistics University, Hamburg, Germany. joe@meiss.com

June 12, 2011

Abstract

Consider a risk-averse decision maker in the setting of a single-leg dynamic revenue management problem with revenue controlled by limiting capacity for a fixed set of prices. Instead of focussing on maximizing the expected revenue, the decision maker has the main objective of minimizing the risk of failing to achieve a given target revenue.

Interpreting the revenue management problem in the framework of finite Markov decision processes, we augment the state space of the risk-neutral problem definition and change the objective function to the probability of failing a certain specified target revenue. This enables us to obtain a dynamic programming solution which generates the policy minimizing the risk of not attaining this target revenue. We compare this solution with recently proposed risk-sensitive policies in a numerical study and discuss advantages and limitations.

Keywords: capacity control, revenue management, multi-period, risk, target level criterion

Please reference this paper as:

Matthias Koenig, Joern Meissner. Risk Minimizing Strategies for Revenue Management Problems with Target Values. Working Paper (available at <http://www.meiss.com>), Lancaster University Management School, 2011.

BIB_TE_X and plain text references are available for download here:

<http://www.meiss.com/en/publications/risk-revenue-management-target-values.html>

1 Introduction

Revenue management systems have become a standard tool in various industries beyond the original airline industry. These newer industries range from cruise lines, rental cars, and media advertising to medical services and event management, see e.g. Talluri and van Ryzin (2005) and Chiang et al. (2007).

We consider a typical revenue management model: a firm operating in a monopolistic setting offering multiple products. These products consume a fixed resource of a limited capacity. The firm sells the products over a finite time horizon. At the end of this time, the salvage value of the resource is assumed to be zero.

The firm can influence its revenue stream by allocating capacity to different classes of demand. Its objective is to find a policy which optimises an objective function. Normally, this objective function is risk-neutral, and the policy is chosen to maximise expected revenue. Such a risk-neutral objective can be motivated by the law of large numbers if the revenue process repeats itself very often, e.g., a daily operating airline flight connection.

However, a risk-neutral policy might not be requested under all scenarios and a risk-averse policy might be advantageous for the decision maker. Lancaster (2003) remarks that a risk-neutral model is often not sufficient, even in the airline industry, as a stable revenue might be preferable due to financial constraints.

In practice, decision makers present some level of risk aversion in revenue management, as mentioned by Bitran and Caldentey (2003). Weatherford (2004) reports the same experience. He observed that airline analysts feel uncomfortable with recommendations of their (risk-neutral) revenue management systems, in particular while waiting for the high-fare passengers a few days before flight departure.

Levin et al. (2008) provide arguments to employ risk-averse policies in scenarios with only a small number of reiterations. They consider an event promoter who has high fixed costs which must be recovered in order to prevent a possible loss. Minimum targets cannot be ignored in such a case. They present further scenarios which require strategic and financial circumstances which necessitate risk considerations.

Taking into account costs of price changes, Koenig and Meissner (2008) illustrate that risk considerations can make a difference when choosing a revenue management strategy.

In recent papers by Barz and Waldmann (2007), Huang and Chang (2009) and Koenig and Meissner (2009), risk-neutral and risk-sensitive policies are analysed. The results show that an appropriate risk-averse policy can be selected if the decision maker knows the parameters representing his level of risk aversion. Such parameters have to be de-

terminated, which is something that is not straightforward in either of the published approaches, whether the underlying concept is an exponential utility function or a discount factor relaxing an optimality condition. Usually, the parameters have to be estimated by running numerical experiments and evaluating risk measures, such as mean-variance or conditional-value-at-risk, on the results.

Thus, we propose using the target-percentile risk measure, discussed by Boda and Filar (2006), as the object function. The target-percentile risk measure computes the probability of the return failing to achieve a previous given fixed target. There are several advantages of using this measure.

First, one important structural property is its time consistency. It says that optimality of decisions should only consider the future. Time consistency is a desirable property for multi-period risk measures as it allows its use in dynamic programming, as shown by the works of Boda and Filar (2006), Artzner et al. (2007), Shapiro (2009). Second, it does not assume a special kind of revenue distribution, as it measures the percentile of the given target level. Third, it is easily interpreted by practitioners and does not require a risk sensitivity parameter which might be difficult to assess. Fourth, numerical computation schemes are available as described by Wu and Lin (1999). Fifth, Boda and Filar (2006) show that multi-stage versions for the well-established risk measure value-at-risk can be developed using the target-percentile measure.

The structure of the paper is as follows. We review the relevant literature in Section 2. In Section 3, we describe our model as a Markov decision process and its extension to apply the target-percentile risk measure. This section also contains some implementation details. Section 4 shows numerical results of our approach and provides a comparison with results of other approaches. Finally, we conclude the paper in Section 5.

2 Related Work

Related work can be divided into three classes: risk and its measurement, in particular multi-period risk measures; Markov decision processes (MDPs) and dynamic programming considering risk in general; revenue management and risk. As literature about risk and its measures is vast, we will only point out some recent publications which we classify to be of interest in context of this paper. We then proceed to discuss links to MDP literature, which provides the basis of our approach.

The expected utility theory as developed by von Neumann and Morgenstern (1947) and the mean-variance measure introduced by Markowitz (1952) are basic works which consider and measure risk. However, both approaches have drawbacks, e.g., utility

theory requires knowledge about one's utility function, and mean-variance theory is only useful with bell-shaped probability distributions and cannot properly be used with every probability distribution. Thus, alternative approaches were investigated. Of these, some risk measures have become more attractive, such as value-at-risk (V@R) and conditional-value-at-risk (CV@R), see Artzner et al. (1999) or Rockafellar and Uryasev (2000). Both measures compute a value for the downside risk with respect to a given confidence value and can be interpreted as kinds of probabilistic constraints. The latter measure has mathematical properties defining its coherence which provide advantages when compared to the former. However, the disadvantage of these measures is they are only suitable for a single-stage decision. They are missing the time-consistency property and, thus, are not suitable for multi-period decision making. This has been recognised, and several authors, such as Boda and Filar (2006), Artzner et al. (2007), and Shapiro (2009), investigate time-consistent risk measures. Such measures are feasible as objective functions for dynamic programming. Hence, they are appropriate for use in revenue management models employing the Bellman equation.

As many problems can be modelled as MDPs (cf. White 1993, Puterman 2005), many authors have investigated risk issues in this context. A good overview about risk in MDPs is provided by White (1988). Thus, we refer to this publication for a more complete overview and point out only briefly the general concepts here. One popular stream of implementing risk in MDPs, which goes back to Howard and Matheson (1972), is using utility functions for computing risk-sensitive policies. Another stream focuses on the variance and probabilistic constraints as risk measures, beginning with the early works of White (1974), Sobel (1982), and ongoing with recent works such as of Sladky and Sitar (2008) and Defourny et al. (2008). We are interested in the target-percentile risk measure which is analysed by Bouakiz and Kebir (1995), Wu and Lin (1999). Their works build the basis for the time-consistent dynamic risk measures proposed by Boda and Filar (2006). The main idea is the extension of the state space in order to use one, or possibly more, variables to keep track of the history.

Most revenue management models use a risk-neutral objective function. We refer to the work of Talluri and van Ryzin (2005) for an overview of these kinds of models. In general, revenue management models are categorised often as capacity control model or dynamic pricing. However, Maglaras and Meissner (2006) discuss the similarities between both categories and give a common formulation in a risk-neutral setting.

The risk-neutral model of dynamic capacity control, which we consider here, was introduced by Lee and Hersh (1993). The corresponding Markov decision process is described by Lautenbacher and Stidham (1999).

The approaches for incorporating risk in revenue management models are analogous to the general decision making under risk: expected utility theory, mean-variance considerations, probabilistic constraints.

Expected utility theory as an element for reflecting risk in revenue management is recommended by Weatherford (2004). He states that the assumption of risk neutrality is not given for many practical scenarios and proposes expected utility theory as a risk-averse solution. Instead the well-adopted (risk-neutral) expected marginal seat revenue model (EMSR), standard algorithms introduced by Beloba (1989), the expected marginal seat utility (EMSU) heuristic can reflect risk sensitivity for decision making.

Recent works of Barz and Waldmann (2007) and Feng and Xiao (2008) are employing expected utility theory, too. Both papers support the application of an exponential utility function to account for risk aversion. Barz and Waldmann (2007) use the Markov decision process formulation of static and dynamic capacity control models, whereas Feng and Xiao (2008) provides closed form solutions from a more general point of view.

Finally, Lim and Shanthikumar (2007) show the equivalence between robust risk-sensitive control and an exponential utility function to dynamic pricing in a single-product setting.

As the first revenue management model with risk considerations, the model of Feng and Xiao (1999) uses variance as its risk measure; in particular, the variance of sales due to price changes. In order to integrate risk into their objective function, they combine expected revenue with a weighted penalty function for the sales variance. The risk sensitivity of the decision maker can be adjusted by the weighting.

With a revenue management application in traffic and networks, Mitra and Wang (2005) compare mean-variance, mean-standard-deviation and mean-CV@R for building an objective function. Their final choice is a mean-standard-deviation formulation. For a truncated Gaussian demand distribution, they demonstrate the impact of risk sensitivity by an efficient frontier.

In an application for hotel revenue management, Lai and Ng (2005) formulate a robust optimisation model for mean versus average deviation.

Recently, Huang and Chang (2009) presented a risk-sensitive modification of the optimality condition for the dynamic capacity control model and investigated their method by measuring mean versus standard deviation in simulation runs. They offer a ranking of their risk-sensitive policies using a Sharpe ratio of revenue and standard deviation.

Illustrating the vulnerability of risk-neutral revenue management due to demand forecast inaccuracy, Lancaster (2003) recommends a relative revenue-per-available-seat-mile-at-risk metric, which integrates risk measurement with the V@R metric. This

metric is the expected maximum of underperformance over a time horizon at a chosen confidence level.

Using both risk measures standard deviation and $CV@R$, Koenig and Meissner (2008) compare the suitability of two different pricing strategies considering the cost of price changes. In a further paper, Koenig and Meissner (2009) evaluate a range of risk-sensitive policies for the dynamic capacity control model.

Risk sensitivity is incorporated by Levin et al. (2008) into a dynamic pricing model of perishable products. Their objective function consists of maximum expected revenue constrained by a desired minimum level of revenue with minimum acceptable probability. This constraint is similar to a $V@R$ formulation. Levin et al. (2008) formulate a hybrid objective function which combines the risk-neutral objective of expected revenue and a penalty term representing risk aversion. Principally, they approach this risk-adjusted maximisation problem by using a further state which a risk-neutral dynamic pricing model does not require. This state keeps track of already gained revenue.

In a capacity control setting, we based our risk incorporation on a state space expansion, too, but our underlying model is derived from a Markov decision process formulation.

3 Description of Model

In the following, we describe the dynamic capacity control problem as a Markov decision process in a similar way as previously done by Lautenbacher and Stidham (1999) and Barz and Waldmann (2007). This model is then expanded in the state space in order to become a model which allows the application of a risk-minimising policy. To achieve this, we follow the approach of Wu and Lin (1999). Our objective function is the target-percentile dynamic risk measure. Finally, we point out some aspects for implementation of this approach.

3.1 Markov Decision Process for Dynamic Capacity Control Model

We consider the capacity control model stated by Lee and Hersh (1993), which is often referred to as dynamic capacity control. Although originally developed for airline revenue management, it can be transferred to other industries. We describe the model in terms of its original airline revenue management context in order to be more intuitive.

We assume that the booking requests follow a Poisson arrival process. Thus, the booking period for a single-leg flight is separated into N decision periods, in such a way that the probability of more than one request can be ignored. The decision periods

are denoted by $n \in \{0, \dots, N\}$. The departure is at $n = 0$. Further, there are k booking classes with fares F_i , $F_1 > F_2 > \dots > F_k$ and $F = \{F_1, \dots, F_k\}$. The probability of a request for fare class i in decision period n is given by $p_{n,i}^r$. Further, we set the probabilities for $n = 0$ to zero for all fare classes: $p_{0,i}^r = 0$; this step just supports our model setting as the last decision will be made at time $n = 1$. The probability of no request at all is $p_{n,0}^r = 1 - \sum_{i=1}^k p_{n,i}^r$. The initial capacity of seats is given by C . The remaining seats are given by $c \leq C$ in a time period.

We have a finite-state, discrete-time, Markov decision process $\Gamma = (S, A, R, P)$ with countable state space S and action space A . Further, R denotes the reward set and P , the set of transition probabilities. Time runs in discrete steps and represents the remaining time before flight departure.

The state space S contains all possible configurations of remaining capacity c and request for a fare class i . Thus $S = \{0, 1, \dots, C\} \times \{0, 1, \dots, k\}$ and a state $(c, i) \in S$ says that we have c seats left and a request for fare class i . We set the fare class 0 with fare $F_0 = 0$, as is often common practice.

Our action space $A(c, i)$ corresponds to the ‘reject’ and ‘accept’ decisions for a given state. We have $A(c, i) = \{0, 1\} \forall (c, i) \in S | c, i > 0$ and $A(c, 0) = A(0, i) = \{0\}$ to only allow the accepting and rejecting of seats at the valid fare prices and not for the artificial class $i = 0$. Overbooking is not allowed.

Let R be the set of rewards (fares) when accepting one booking. Rewards are denoted by $r_n(s, a) \in R$ with $s \in S, a \in A$ and $r_n((c, i), a) = aF_i$ for $n, c > 0$ and zero otherwise. The transition probabilities $p \in P$ are defined for states $(c, i), (c - a, j) \in S$ with $a \in A$ by $p_n((c - a, j) | (c, i), a) = p_{n,j}^r$ for $n = N, N - 1, \dots, 0$, and 0 otherwise.

A decision maker decides on a sequence of rules $a_n = d_n(c_n, i_n)$, which determine a policy $\pi = \{d_n, d_{n-1}, \dots, d_1\}$. Thus, one of the possible actions is chosen by accepting or rejecting a booking request for each state (c_n, i_n) .

Now let $\rho_N^\pi(c, i) = \sum_{n=0}^N r_n$ denote the random variable of the gained revenue for a particular policy π beginning with capacity c and request i at N remaining time steps. Its expected revenue is given by

$$v_N^\pi(c, i) = \mathcal{E}_\pi [\rho_N^\pi(c, i)] = \mathcal{E}_\pi \left[\sum_{n=1}^N r_n((c_n, i_n), d_n(c_n, i_n)) \right].$$

The maximal expected revenue and its associated risk-neutral policy can be computed by the Bellman equation for this problem. However, we are interested in a policy which minimises the time-consistent dynamic risk measure of not achieving a target revenue x in the accumulated return.

3.2 Expanded Markov Decision Process for Minimising Risk of Failing Target

We are interested in minimising the risk of not attaining a specified target revenue x_N for the dynamic capacity control model. Thus, we want to find a policy π which minimises the objective function representing the probability of not achieving a previous specified target level x_N . In order to derive this objective function, we follow the approaches mentioned by White (1988), Wu and Lin (1999) and Boda and Filar (2006) and expand the Markov decision process Γ by a larger state space. The extended Markov decision process $\tilde{\Gamma}$ is similar to Γ . It consists of $\tilde{\Gamma} = (\tilde{S}, \tilde{A}, \tilde{R}, \tilde{P}) = (\tilde{S}, A, R, P)$, as described below.

The state space S is replaced by the new state space $\tilde{S} = S \times \mathbb{R}$ with elements $((c, i), x)$ with $\mathbb{R} = (-\infty, \infty)$. The new state space \tilde{S} consists of states of the configurations of remaining capacity c and a request for fare class i , and additionally, a revenue target x . All state variables are updated over time, e.g., the revenue target x decreases by the realised fare price in accordance with decrementing c by selling a seat.

The action space \tilde{A} is generated from action state A by $\tilde{A}((c, i), x) = A(c, i)$, $\forall ((c, i), x)$ and, thus, $\tilde{A} = \cup_{((c, i), x) \in \tilde{S}} \tilde{A}((c, i), x) = \cup_{(c, i) \in S} A(i) = A$. In a similar way, the reward set \tilde{R} is build from R . For $\tilde{s} \in \tilde{S}$, $a \in A$, the reward $\tilde{r}_n(\tilde{s}, a) \in \tilde{R}$ is $\tilde{r}_n((c, i), x, a) = aF_i$ for $c, i > 0$ and zero otherwise. Thus, $\tilde{R} = R$.

As well, $\tilde{P} = P$, as the transition probabilities \tilde{P} are determined by P . We have $\tilde{p} \in \tilde{P}$ and, with states $((c, i), x), ((c - a, j), x - aF_i) \in \tilde{S}$ and $a \in A$, the transition probability is given by $\tilde{p}_n((c - a, j), x - aF_i) | ((c, i), x), a) = p_{n,j}^r$ for $n = N, N - 1, \dots, 0$ and else 0.

We are interested in the probability that our obtained total revenue does not attain a target level x_N . Let the set of deterministic Markovian policies be $\tilde{\Pi}$ and let the random variable for the cumulative gained reward, applying policy $\tilde{\pi} \in \tilde{\Pi}$ beginning with capacity c , request i , remaining time steps N , and initial target x_N , be $\rho_N^{\tilde{\pi}}((c, i), x_N) = \sum_{n=0}^N \tilde{r}_n$. For the policy $\tilde{\pi}$, the target-percentile risk measure is defined as

$$V_N^{\tilde{\pi}}((c, i), x_N) := \mathcal{P} \left(\rho_N^{\tilde{\pi}}((c, i), x_N) \leq x_N \right), \quad (1)$$

where \mathcal{P} denotes a probability. The time consistency property of the target-percentile risk measure can be shown as demonstrated by Boda and Filar (2006).¹

¹We have non-stationary transition probabilities in our model, as opposed to the assumption of Boda and Filar (2006). Nevertheless, their approach can be used. Our non-stationary case could be transformed to a stationary one by a state augmentation with n which would resolve the dependence of the probabilities on n .

Thus, we are looking now for an optimal policy $\tilde{\pi}^*$ for each objective function $V_N^{\tilde{\pi}}((c, i), x_N)$ that minimises the risk of failing target level x_N :

$$\tilde{\pi}^* = \operatorname{argmin}_{\tilde{\pi} \in \tilde{\Pi}} \left\{ V_N^{\tilde{\pi}}((c, i), x_N) \mid ((c, i), x_N) \in \tilde{S}, n \geq 1 \right\}. \quad (2)$$

The associated percentile (minimum risk level for x_N) is denoted $V_N^{\tilde{\pi}^*}$.²

Following Wu and Lin (1999) and Boda and Filar (2006), we can derive the dynamic programming equations for computation of the minimum percentile $V_N^{\tilde{\pi}^*}$.³ We attain the following equations for $x \in \mathbb{R}, \forall (c, i) \in S$:

$$V_0^{\tilde{\pi}^*}((c, i), x_0) = \begin{cases} 1 & x_0 > 0 \\ 0 & \text{otherwise,} \end{cases}$$

$$V_n^{\tilde{\pi}^*}((c, i), x_n) = \min_{a \in A} \left\{ \sum_{j=0}^k p_{n-1,j}^r V_{n-1}^{\tilde{\pi}^*}((c-a, j), x_n - aF_i) \right\}, n \in [1 \dots N]. \quad (3)$$

The initial probabilities are one for a target $x_0 > 0$ in time $n = 0$ as there is no remaining time for earning any value. Otherwise, every target $x_0 \leq 0$ will definitely be met. As the initial revenue is zero and the fare prices are positive, our revenue will be always be greater than or equal to zero. Thus, the probability that we do not attain a target $x_0 \leq 0$ is zero, and the probability that we do not attain a target $x_0 > 0$ is one.

Note that the final percentile of all time periods is determined by $V_{N+1}^{\tilde{\pi}^*}(c, i, x_N)$; in $V_N^{\tilde{\pi}^*}(c, i, x_N)$, we already know the requested class i at time N .

The optimal policy $\tilde{\pi}^*$ can be computed from the minimum percentile $V_N^{\tilde{\pi}^*}$ by Equation 2 for a given target level x_N . It should be pointed out that an optimal policy describes one way to obtain the target level, but several optimal policies might exist. Therefore, if more than one decision rule can be chosen in a certain state in order to achieve the minimum percentile we select the decision rule which contributes most to the revenue. In particular, we prefer to accept a request if the probabilities of both possible decisions $a \in \{0, 1\}$ are equal when determining the minimum in Equation 3, $\sum_{j=0}^k p_{n-1,j}^r V_{n-1}^{\tilde{\pi}^*}((c, j), x) = \sum_{j=0}^k p_{n-1,j}^r V_{n-1}^{\tilde{\pi}^*}((c-1, j), x - F_i)$ and the risk-neutral solution would accept the request, too. In this manner, we achieve the same probabilities regarding the target but with the policy which yields the greater expected revenue.

Furthermore, if we have in some state achieved the target level the following states can be arbitrarily chosen. In practice, the policy for the ongoing states should be optimised then under another criterion, such as the expected revenue. Moreover, if the

²Actually, the percentile is $100 \times V_N^{\tilde{\pi}^*}$, but we follow the denotations of Boda and Filar (2006).

³see Appendix.

target can never be obtained in the given setting, all policies are equally improper and no optimal target-percentile policy exists. In both cases, we apply the risk-neutral policy which maximises the expected revenue.

Example

In order to illustrate the method, we give a stylised example. Consider only two classes with fares $F_1 = 200; F_2 = 100$, two remaining time periods $N = 2$, only one seat left $C = 1$, and the probabilities for arrivals $p_{1,1} = 0.10, p_{1,2} = 0.15, p_{2,1} = p_{2,2} = 0.20$. Thus, for example, the probability of a request of fare 2 in period 1 before departure is 15 percent. We have only a few scenarios in this setting: if a request for a distinct fare class comes in period 2 before departure, we can accept it or reject this fare class and then wait for possible arrivals in the last period and, if they appear, accept. It is easy to see that the policy which always accepts (expected revenue of 81) is better off when compared with others. However, consider that now we want the best policy for a target value of 200. The expected revenue maximising policy fails that target with probability of 0.74. A better choice for this target would be only acceptance of the highest fare class, a policy which fails only with a likelihood of 0.72. The computation using the proposed method is shown in Figure 1. Note that the transition probabilities of the last period $n = 0$ are zero for all classes except for the artificial class 0, where they are one.

Implementation Details

The dynamic programming formulation given in Equation 3 is inefficient for implementation. We can point out two approaches to obtain a better implementation.

First, we can apply an often used transformation of the dynamic programming formulation, allowing us to scale down the state space. Thus, introducing the operator $T_n(c, x) := \sum_{i=0}^k p_{n,i}^r V_n(c, i, x)$ helps by reducing the state space by variables representing the fare class of an arrival. Defining $W_n(c, x) := T_n(c, x)V_n(c, i, x)$, we transform Equation 3, as follows, for $x \in \mathcal{R}, c \in \{0, \dots, C\}$:

$$\begin{aligned}
 W_0^{\tilde{\pi}^*}(c, x_0) &= \begin{cases} 1 & x_0 > 0 \\ 0 & \text{otherwise,} \end{cases} \\
 W_n^{\tilde{\pi}^*}(c, x_n) &= T_n(c, x_n)V_n^{\tilde{\pi}^*}(c, i, x_n) \\
 &= \sum_{i=0}^k p_{n,i}^r \min_{a \in A} \{W_{n-1}^{\tilde{\pi}^*}(c - a, x_n - aF_i)\}. \tag{4}
 \end{aligned}$$

$$\begin{aligned}
 V_0(c, i, x) &= 1, x > 0, \\
 V_0(c, i, x) &= 0, x \leq 0 \\
 V_1(0, i, x) &= 1 \cdot V_0(0, 0, x) + 0 \cdot V_0(0, 1, x) + 0 \cdot V_0(0, 2, x) \\
 &= 1, x > 0, \\
 V_1(0, i, x) &= 1 \cdot V_0(0, 0, x) + 0 \cdot V_0(0, 1, x) + 0 \cdot V_0(0, 2, x) \\
 &= 0, x \leq 0 \\
 V_1(1, 0, 200) &= 1 \cdot V_0(1, 0, 200) + 0 \cdot V_0(1, 1, 200) + 0 \cdot V_0(1, 2, 200) \\
 &= 1 \\
 V_1(1, 1, 200) &= \min\{1 \cdot V_0(1, 0, 200) + 0 \cdot V_0(1, 1, 200) + 0 \cdot V_0(1, 2, 200), \\
 &\quad 1 \cdot V_0(0, 0, 0) + 0 \cdot V_0(0, 1, 0) + 0 \cdot V_0(0, 2, 0)\} \\
 &= 0 \\
 V_1(1, 2, 200) &= \min\{1 \cdot V_0(1, 0, 200) + 0 \cdot V_0(1, 1, 200) + 0 \cdot V_0(1, 2, 200), \\
 &\quad 1 \cdot V_0(0, 0, 100) + 0 \cdot V_0(0, 1, 100) + 0 \cdot V_0(0, 2, 100)\} \\
 &= 1 \\
 V_1(1, 0, 100) &= 1 \cdot V_0(1, 0, 100) + 0 \cdot V_0(1, 1, 100) + 0 \cdot V_0(1, 2, 100) \\
 &= 1 \\
 V_1(1, 1, 100) &= \min\{1 \cdot V_0(1, 0, 100) + 0 \cdot V_0(1, 1, 100) + 0 \cdot V_0(1, 2, 100), \\
 &\quad 1 \cdot V_0(0, 0, -100) + 0 \cdot V_0(0, 1, -100) + 0 \cdot V_0(0, 2, -100)\} \\
 &= 0 \\
 V_1(1, 2, 100) &= \min\{1 \cdot V_0(1, 0, 100) + 0 \cdot V_0(1, 1, 100) + 0 \cdot V_0(1, 2, 100), \\
 &\quad 1 \cdot V_0(0, 0, 0) + 0 \cdot V_0(0, 1, 0) + 0 \cdot V_0(0, 2, 0)\} \\
 &= 0 \\
 V_2(1, 0, 200) &= p_{1,0}V_1(1, 0, 200) + p_{1,1}V_1(1, 1, 200) + p_{1,2}V_1(1, 2, 200) \\
 &= 0.90 \\
 V_2(1, 1, 200) &= \min\{p_{1,0}V_1(1, 0, 200) + p_{1,1}V_1(1, 1, 200) + p_{1,2}V_1(1, 2, 200), \\
 &\quad p_{1,0}V_1(0, 0, 0) + p_{1,1}V_1(0, 1, 0) + p_{1,2}V_1(0, 2, 0)\} \\
 &= 0 \\
 V_2(1, 2, 200) &= \min\{p_{1,0}V_1(1, 0, 200) + p_{1,1}V_1(1, 1, 200) + p_{1,2}V_1(1, 2, 200), \\
 &\quad p_{1,0}V_1(0, 0, 100) + p_{1,1}V_1(0, 1, 100) + p_{1,2}V_1(0, 2, 100)\} \\
 &= 0.90 \\
 V_3(1, 0, 200) &= p_{2,0}V_2(1, 0, 200) + p_{2,1}V_2(1, 1, 200) + p_{2,2}V_2(1, 2, 200) \\
 &= 0.72
 \end{aligned}$$

 Figure 1: Exemplary computation of $V_N^{\tilde{\pi}^*}(c, i, x)$, superscript omitted.

i	F_i	$1 \leq n \leq 4$	$5 \leq n \leq 11$	$12 \leq n \leq 18$	$19 \leq n \leq 25$	$26 \leq n \leq 30$
1	200	0.15	0.14	0.10	0.06	0.08
2	150	0.15	0.14	0.10	0.06	0.08
3	120	0	0.16	0.10	0.14	0.14
4	80	0	0.16	0.10	0.14	0.14

 Table 1: Fares and request probabilities for fare class i and time period n .

Second, the computation of all possible cumulative rewards given by the variable x might be very impractical and could be done on a suitable grid for larger problems, as described in the works of Wu and Lin (1999) and Boda et al. (2004). We refer to the latter publication for a deeper analysis of this approximative dynamic programming solution. Instead of taking into account every value which x can take while solving the problem, we can use an interval between 0 and the initial x_N . E.g., if we have $m + 1$ values for this interval, the interval could be $[\mathcal{Y}_0, \mathcal{Y}_1, \mathcal{Y}_2, \dots, \mathcal{Y}_m] = [0, \frac{x_N}{m}, \frac{2x_N}{m}, \dots, x_N]$. The computation of $W_n^{\tilde{\pi}^*}$ proceeds now only on these samples, and we obtain a grid of values $\{(\mathcal{Y}_0, W_n^{\tilde{\pi}^*}(c, \mathcal{Y}_0)), \dots, (\mathcal{Y}_m, W_n^{\tilde{\pi}^*}(c, \mathcal{Y}_m))\}$ for $c \in \{0, \dots, C\}$. The program fills the values on the grid and rounds occurring values to the upper value: $W_n^{\tilde{\pi}^*}(c, \mathcal{Y}) = W_n^{\tilde{\pi}^*}(c, \mathcal{Y}_{j+1}) \forall \mathcal{Y} \in (\mathcal{Y}_j, \mathcal{Y}_{j+1}]$. Alternative rounding schemes could be used.

4 Numerical Simulation and Results

In their introductory paper about dynamic capacity control, Lee and Hersh (1993) used an example which also served for illustration in the recent papers of Barz and Waldmann (2007), Huang and Chang (2009) and Koenig and Meissner (2009). Thus, we can also demonstrate the proposed target-percentile policy in the same exemplary setup.

Exemplary Simulation Setup

There are $N = 30$ time periods before departure, and the initial number of seats is $C = 10$. The four fare classes are $F_1 = 200, F_2 = 150, F_3 = 120, F_4 = 80$. The probabilities for a request of a fare class in a given time period are shown in Table 1.

In order to see how the target-percentile policy works, we conducted an experiment with 1000 sample runs. Random arrivals were simulated in a Monte Carlo manner using the values of Table 1. When compared with other proposed policies, the same sample paths (random arrivals) were used, of course.

A single simulation run is initialised with values for remaining seats, time periods before departure, and a policy. The policy contains for each state the acceptable fare classes. The state is described by remaining time periods, remaining seats, and a remaining target value. The simulation then continues to loop over the time periods until the departure time zero is reached. Inside the loop, a random generator simulates requests for fare classes which are accepted if the current policy allows acceptance of the class or else rejected. An update of the state is as follows: time periods are always decremented by one, seats are decremented only if a fare is accepted, and the target value is decremented by the gained fare price. Our policies have one more dimension than the policies of the above mentioned references. Hence, we can abandon the target-percentile dimension when simulating the policies used for comparison.

Policy Illustration

Figure 2 visualises the policy $\tilde{\pi}^*$ for the described example. We see slices through a three-dimensional matrix codifying protection levels in colour. Basically, this three-dimensional matrix displays the protection levels — the maximum allowed fare class — for each state (c, x) in time n with initial target of 1200. In order to use the policy, we start in the state $(30, 1200)$ with 30 time periods to go. This is the top corner on the right hand side of the presented box. The state at this position in the matrix has a protection level which lets one decide how to act at this point in time before departure. Only fare classes with higher prices than the associated protection level are accepted. As time marches on, one moves always one step further along the time dimension to departure time zero; this is parallel to the south-west direction in the figure. The policy decides now which way to move in both other dimensions. An acceptance of a request causes a move downwards along the dimension of the capacity, orthogonally downwards in the matrix. Finally, the price of an accepted fare dictates how to move in the target direction, along the north-west direction in the figure. Thus, considering the figure, the simulation will generate random trajectories from the top corner on the right hand side to the bottom corner on the left hand side. Of course, the end of each trajectory will often be different due to the random realisations but it has to end with coordinate $n = 0$.

We illustrate the effect of changing a target level on the policy in Figure 3. The figure shows for three different target levels (1200, 1400, 1550) the corresponding policies when revenue has not yet been gained. The effect of increasing the target can be observed by the right hand side of the shown protection level matrices. E.g., a capacity of at least 6 seats is required for a target of 1200, but a capacity of at least 7 seats is required for a target of 1400. Only the highest fare class is accepted when only 6 seats

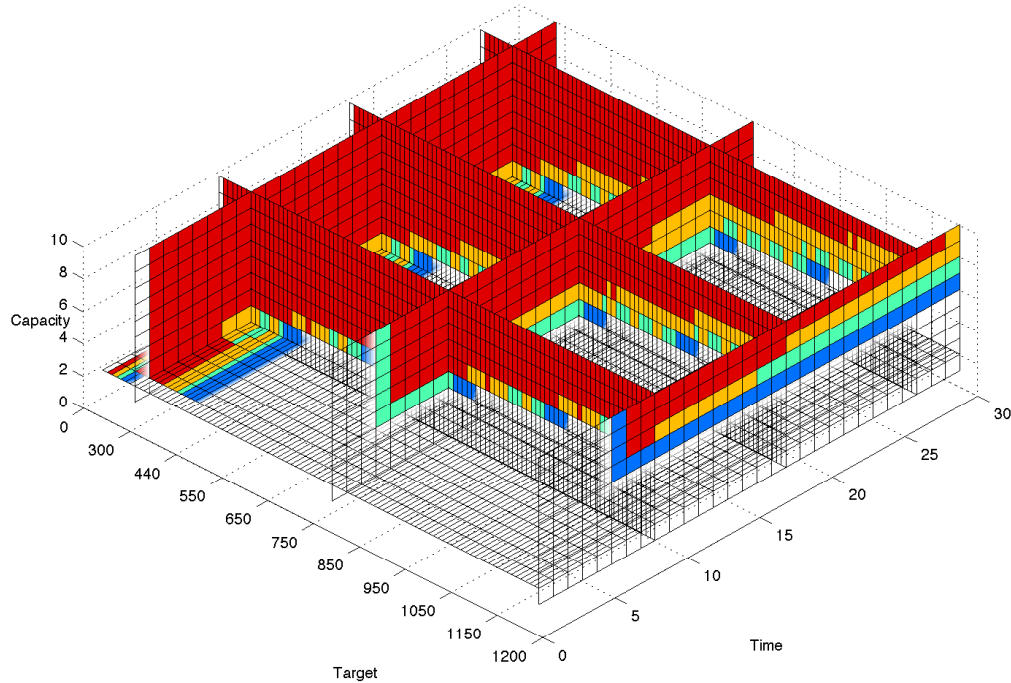


Figure 2: Target level policy : Protection levels of policy $\tilde{\pi}^*$ are shown for each state (c, x) with remaining capacity $0 \leq c \leq 10$, revenue target $0 \leq x \leq 1200$ in time $n, 1 \leq n \leq 30$. The protection levels for the four fare classes are visualised by colour: class 1 (blue), class 2 (green), class 3 (orange), and class 4 (red). Only fare classes of higher price than indicated by the colour are accepted in a state. Note, states which offer no optimal solution because the revenue is not obtainable, given the remaining capacity and time, are displayed in transparency.

are available for the target of 1200, respectively, when only 7 seats are available for the target of 1400. In case of attaining the 1550 target, 7 seats are required but the two highest fare classes are acceptable then. Figure 3 illustrates the pure target policies. As already mentioned, we apply the risk-neutral policy in states which do not allow to achieve a target; this is not shown in this figure.

Evaluation

As the proposed policy optimises the target-percentile, we start our evaluation with different (obtainable) target revenues, comparing the theoretical and the simulation results. As mentioned in Section 3.2, there are scenarios when target revenue is achieved but time is remaining and one or more seats are left. We present the average of remain-

Target value	1600	1500	1400	1300	1200
$V_n^{\tilde{\pi}^*}$	0.740	0.528	0.336	0.183	0.088
Simulation of target value policies					
Failed target ($\times 100\%$)	0.749	0.510	0.354	0.189	0.117
Revenue	1314.8	1362.5	1367.4	1345.8	1319.9
Std. deviation	276.97	242.76	191.56	161.23	160.81
Remaining time	0.04	0.22	0.82	1.93	3.87
Remaining seats	0.01	0.04	0.11	0.23	0.49
Simulation of risk-neutral policy					
Failed target ($\times 100\%$)	0.821	0.621	0.420	0.271	0.158
Revenue			$\leftarrow 1402.2 \rightarrow$		
Std. deviation			$\leftarrow 210.36 \rightarrow$		

Table 2: Results of policy simulation for different target values. The probabilities and averages for failing to achieve a target are given (lower means better) and also, averages of remaining time and seats if target could be achieved. An expected revenue optimising risk-neutral policy yields theoretically 1407.2. For comparison, the results of the simulated risk-neutral policy are given.

ing time and seats for such cases as well. Further, the averaged revenue is computed by switching to the risk-neutral policy when the target has been achieved. Table 2 shows the results for five different targets. The average of failed cases in the simulation is plausible within numerical errors to the theoretical percentile, validating that the policy does as expected.

The expected revenue for the analysed problem is 1407.2. Looking at the results of Table 2, we see that a policy which aims towards a lower target revenue than the expected value accepts an upcoming request early in time. Decisions are made soon and not postponed to later periods. This effect is easily observable as remaining time and seats decrease, while the target is increasing. Of course, policies with lower targets have a greater probability for reaching the target. It can be more easily obtained by accepting requests early, thus leaving more time for balancing against having no profitable requests in the next time periods.

The average revenues of the target policies are in each case lower than that of the risk-neutral policy. The standard deviation of these revenues grows with an increasing target, although when compared with the risk-neutral case, their policies less often fail the targets. This can be explained by comparing the distribution histograms of the revenues of the policies.

Figure 4 shows the distribution histograms of 1000 simulation runs of three policies: one with low target 1200, one with high target 1400, and a risk-neutral one maximising expected revenue.

The distribution associated with the low target has its peak above its target value 1200 and a slight negative skew. It has only small frequencies for values lower than 1200 but also for values higher than 1500, as its standard deviation from Table 2 also emphasises. It has a peak at 1300.

The risk-neutral solution shows a negative skewed distribution with a peak at 1500 with a long tail to very low values, though some high revenues at 1800. Compared with the policy with target 1200, its revenues are more often below 1200; however, given the revenue is greater 1200, it will be better off. Its risk of falling below 1200 remains higher than the risk of the low target policy.

The distribution of the policy with high target 1400 has a strong negative skew with a long tail to low values, too. The peak of the distribution is at 1400. Compared with the risk-neutral counterpart, this policy shifts frequency from 1300 to 1400 revenue. The target is achieved mainly at the expense of 1300 revenue and greater than 1500 revenue. Further, it shows also higher frequencies for low revenue than both other policies. Hence, if it fails the target, there is a greater risk of obtaining only low revenue.

The histogram demonstrates that the policy with low target aims at a lower average revenue and smaller variance, but the policy with a higher target, near to the expected revenue of the risk-neutral solution, does not.

In order to evaluate the performance of target revenue policies in more detail, we compare them with the risk-sensitive policies derived from expected utility theory, as in Barz and Waldmann (2007). We select the latter policies for comparison as they result from optimising the dynamic capacity control model using an exponential utility⁴ and no heuristics. Referring to the recent works of Huang and Chang (2009) and Koenig and Meissner (2009), we view the mean, standard deviation, and CV@R of the policies. The CV@R is a measure for the expected revenue given the revenue is below a certain quantile specified by a confidence level α ; it is the expected value in the α percent of worst cases.

Table 3 compares both types of risk-sensitive policies. Beyond the mean, standard deviation, and CV@R with confidence level 5%, the observed relative frequency of failing the 1000 target is given. We see that the target policy for 1000 has the least risk of failing it. However, it is also observable that the target policies only limit the risk of failing their certain target and do not provide more preferable results in terms of the

⁴Instead of searching the policy π for $\max_{\pi} \mathcal{E}_{\pi} [\rho_N^{\pi}(c, i)]$, the expected utility approach with exponential utility uses $\max_{\pi} \mathcal{E}_{\pi} [\exp(-\gamma \rho_N^{\pi}(c, i))]$, where γ represents a factor for the level risk aversion.

Policy	mean	std. dev.	CV@R	freq.($\rho < 1000$)
Target 800	1354.2	177.3	941.2	0.028
Target 900	1349.6	171.8	960.4	0.029
Target 1000	1331.6	170.3	959.4	0.020
Target 1100	1322.0	165.5	951.8	0.031
Target 1200	1319.9	160.8	927.0	0.029
Target 1300	1345.8	161.2	910.2	0.041
Target 1400	1367.4	191.6	874.0	0.050
Utility $\gamma = 0.010$	1359.5	162.3	953.2	0.027
Utility $\gamma = 0.005$	1383.8	176.3	939.6	0.029
Utility $\gamma = 0.001$	1402.3	206.8	886.6	0.050

Table 3: Comparison between two risk-sensitive policies: target-percentile optimising and exponential utility function optimising policies (the risk aversion increases in conjunction with γ). CV@R is for $\alpha = 5\%$.

other measures. The expected utility based policies have higher average revenues than most target policies. If the target policies aim at a level greater than 1200, the CV@R drops down with further increasing the level. The CV@R of the policies employing an exponential utility function decreases with increasing level of risk aversion. The standard deviation increases with higher target and higher level of risk sensitivity for both types of policies. Discussed already by Figure 4, the CV@R results also show that the target policies do not limit the risk of obtaining only low revenues in the worst cases. Further, it is interesting that the policies aimed at targets different from 1000 do not guarantee good performance regarding the 1000 target.

This effect becomes more observable in the distribution histogram of the 1000 revenue target policy and the expected utility policy with high risk aversion $\gamma = 0.01$, as shown in Figure 5. The target policy has a lower average revenue, a higher 5% CV@R, and a higher standard deviation than the exponential utility policy, and it achieves at least a revenue of 1000 in more cases. The frequencies for the revenues 800 and 900 are lower for the target policy than for the exponential utility policy. But, the target policy has higher frequencies for revenues between 1000 and 1200 and lower frequencies between 1300 and 1800, with exception of 1700, than its counterpart. This explains the lower mean revenue.

Figures 4 and 5 show that the target policies dent the distribution slightly below the target. Thereby, the whole distribution, for values lower and greater than the target, is influenced. Frequencies below this dent may increase as frequencies for the target do.

In particular, distribution lower than the target need not be modified in a favourable manner regarding the lowest revenues, that is to say the worst cases.

Further Numerical Experiments

We did further numerical experiments beyond the previous illustrative one. In order to investigate the target policy, we show five more scenarios which differed with respect to their load factor. The load factor is given by $\rho = \frac{1}{C} \sum_{n=0}^N \sum_{j=1}^k p_{jn}$ and gives information about demand in relation to capacity. The previous example had a load factor of $\rho = 1.32$.

We changed only the request probabilities of the previous example and hold the other parameters fixed to get further scenarios. To this end, we built the further scenarios by choosing random request probabilities which yielded different load factors. Then we simulated 1000 sample runs with each scenario. Table 4 shows results of the risk-neutral policy and of the target policy which were applied to the scenarios. We selected those targets for each policy which were 115%, 100% and 85% times the expected revenue of the risk-neutral policy.

As Table 4 shows, the target policies achieved the desired target more often than the risk-neutral policies in the numerical simulations. That advantage of the target policies increased along with the increasing load factor.

5 Conclusions

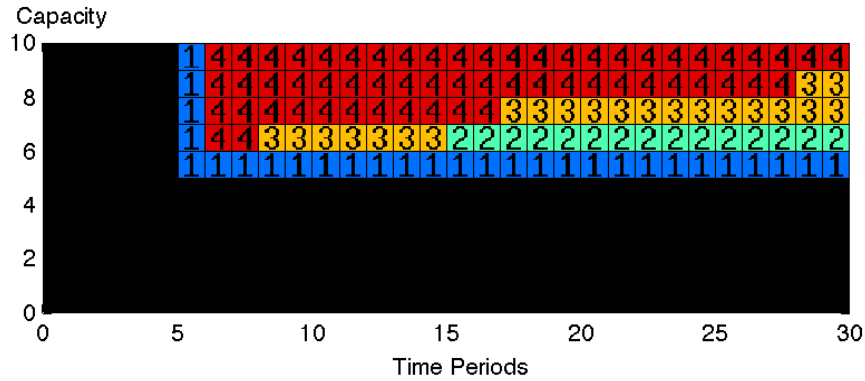
A risk-averse policy minimising the failure of a previously defined, certain revenue target has been proposed for a revenue management problem, namely the dynamic capacity control setting. This policy is derived by extending the state space of the Markov decision process formulation of the problem. We have discussed aspects for implementing the policy numerically.

In numerical experiments, we have analysed the proposed policy and evaluated against risk-neutral and another risk-sensitive policies. We have compared the mean, standard deviation, and conditional-value-of-risk of those policies. The optimal policy for a given target revenue focuses on minimising the likelihood of the failing of this certain target but does not compensate for other risk measures.

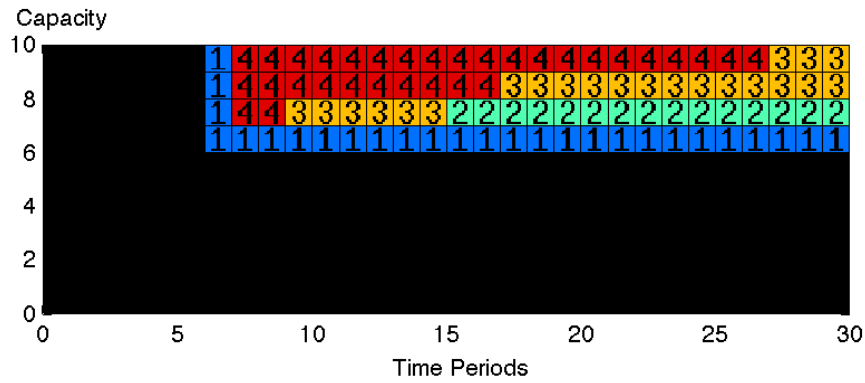
The analysis of the revenue distributions of the target revenue aimed policies in numerical experiments disclose how important correct understanding of such a policy is when applied. The decision maker must be aware of its limitations; in particular, that it is the policy with lowest probability of failing the target, but the probability of worst

outcomes are not eliminated. However, using a low target revenue supports limiting such risk.

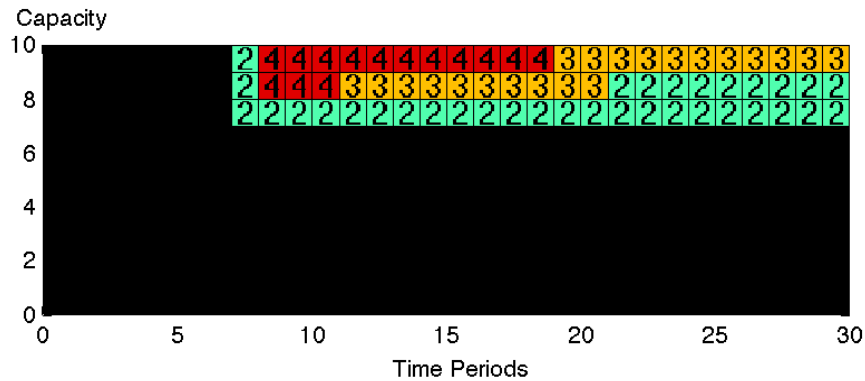
The presented approach can be further developed in order to achieve a policy which optimises value-at-risk as proposed by Boda and Filar (2006). Furthermore, it also offers the basis for the development of investigating policies balancing out mean revenue versus target achievement.



(a) Target $x_N = 1200$



(b) Target $x_N = 1400$



(c) Target $x_N = 1550$

Figure 3: Illustration of the effect by changing the target revenue on the policy: The figures show the protection levels of the policies for a capacity between 1 and 10 seats, a remaining time between 1 and 30 periods and when revenue has not yet been collected. Only fare classes up to the indicated level are accepted in a state. States which do not allow a policy to attain the target are shown in black colour.

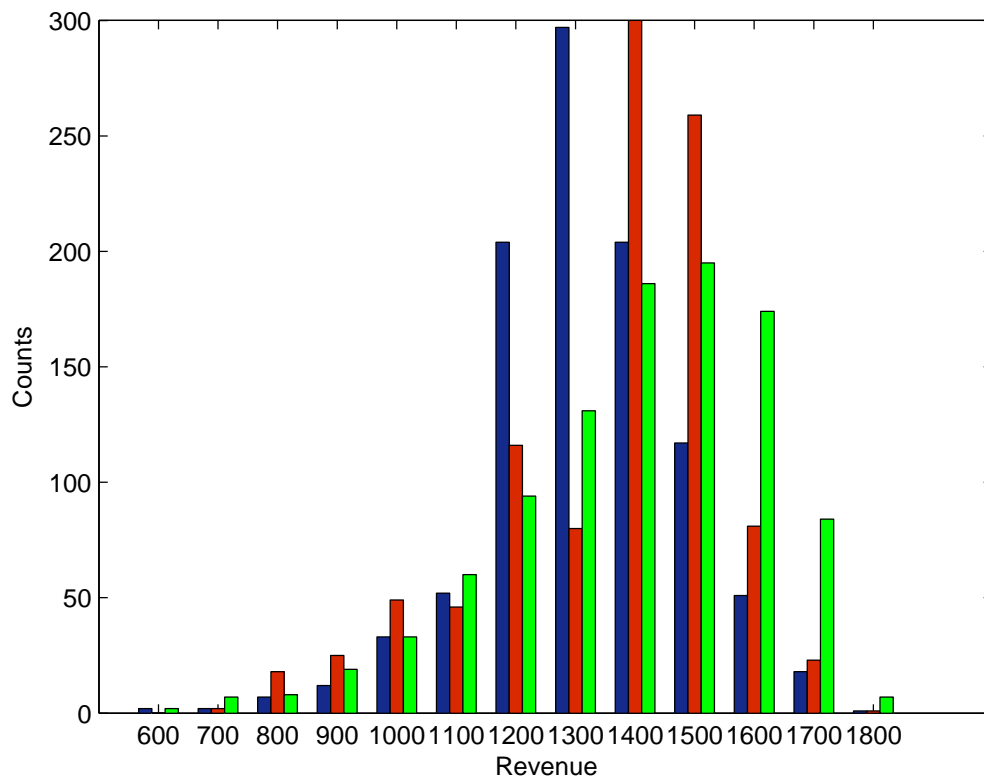


Figure 4: Histogram of the distributions of gained revenue in the simulation using policies with revenue targets 1200 (left blue bars), 1400 (middle red bars) and the risk-neutral policy (right green bars).

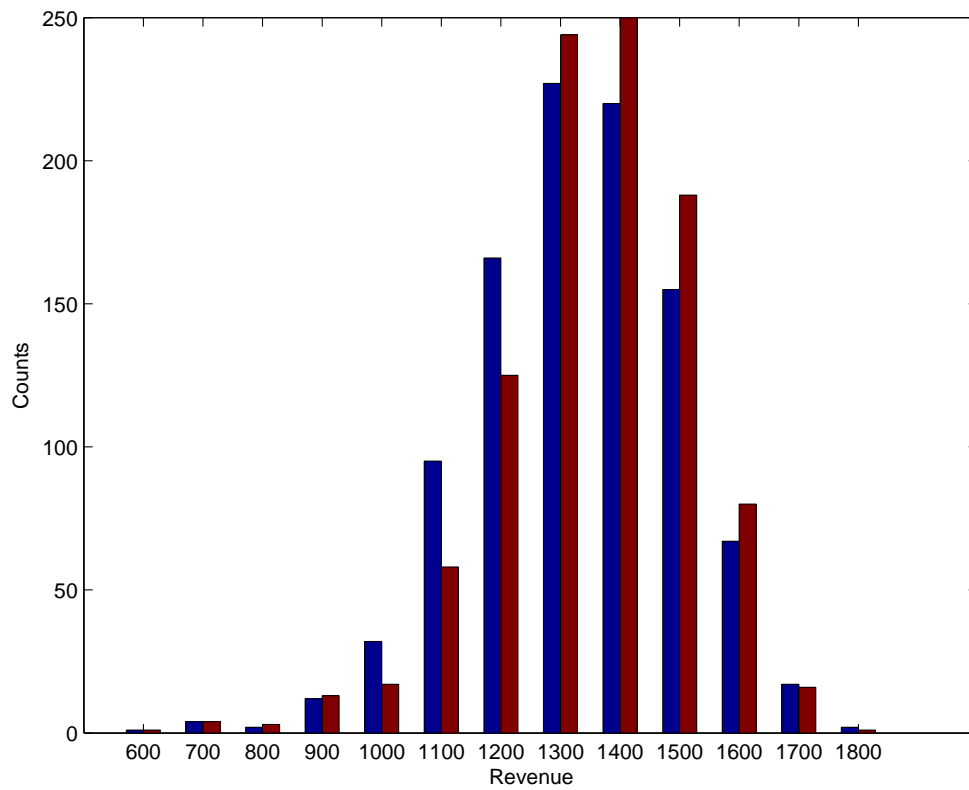


Figure 5: Histogram of the distributions of gained revenue in the simulation. The left blue bars show revenue frequencies gained by target policy with target 1000, the right red bars show frequencies by an expected utility optimising policy using exponential utility function with $\gamma = 0.01$.

$\rho = 0.89$, exp. rev. = 1157.2, sim. rev. = 1155.9				
Target	$V_n^{\tilde{\pi}^*}$	fail. targ.	diff. fail. vs. risk neut.	revenue
1340	0.697	0.699	0.001	1145.0
1160	0.453	0.449	0.022	1151.5
990	0.255	0.258	0.002	1156.6
$\rho = 0.99$, exp. rev. = 1280.1, sim. rev. = 1277.8				
Target	$V_n^{\tilde{\pi}^*}$	fail. targ.	diff. fail. vs. risk neut.	revenue
1480	0.750	0.750	0.038	1251.6
1290	0.453	0.470	0.070	1259.2
1090	0.210	0.212	0.013	1271.6
$\rho = 1.11$, exp. rev. = 1356.3, sim. rev. = 1358.7				
Target	$V_n^{\tilde{\pi}^*}$	fail. targ.	diff. fail. vs. risk neut.	revenue
1560	0.779	0.776	0.072	1320.2
1360	0.398	0.387	0.046	1345.1
1160	0.131	0.119	0.039	1335.8
$\rho = 1.48$, exp. rev. = 1482.0, sim. rev. = 1484.3				
Target	$V_n^{\tilde{\pi}^*}$	fail. targ.	diff. fail. vs. risk neut.	revenue
1710	0.828	0.821	0.073	1391.4
1490	0.386	0.382	0.070	1446.2
1260	0.070	0.075	0.037	1389.6
$\rho = 1.76$, exp. rev. = 1525.5, sim. rev. = 1522.7				
Target	$V_n^{\tilde{\pi}^*}$	fail. targ.	diff. fail. vs. risk neut.	revenue
1760	0.931	0.934	0.034	1269.2
1530	0.333	0.320	0.160	1477.9
1300	0.012	0.014	0.057	1437.4

Table 4: Results of numerical simulation with scenarios which differed by their load factors ρ . Revenues are averaged over 1000 sample runs. The given differences are the observed relative frequencies of failed target instances of the risk-neutral policy minus the observed relative frequencies of failed target instances of the target policy. The differences show how more often the target level policies achieved the target but the risk-neutral ones did not. E.g., in the example of the last row of the table, the target policy did not failed in 5.7% of all sample runs but the risk-neutral one did.

References

- Artzner, Philippe, Freddy Delbaen, Jean-Marc Eber, David Heath. 1999. Coherent measures of risk. *Mathematical Finance* **9** 203–228.
- Artzner, Philippe, Freddy Delbaen, Jean-Marc Eber, David Heath, Hyejin Ku. 2007. Coherent multiperiod risk adjusted values and bellman's principle. *Annals of Operations Research* **152** 5–22.
- Barz, C., K.-H. Waldmann. 2007. Risk-sensitive capacity control in revenue management. *Mathematical Methods of Operations Research* **65**(3) 565–579.
- Beloba, P. P. 1989. Application of an probabilistic decision model to airline seat inventory control. *Operations Research* **37**(2) 183–197.
- Bitran, Gabriel, Rene Caldentey. 2003. An overview of pricing models for revenue management. *Manufacturing & Service Operations Management* **5**(3) 203–229.
- Boda, Kang, Jerzy A. Filar. 2006. Time consistent dynamic risk measures. *Mathematical Methods of Operations Research* **63** 169–186.
- Boda, Kang, Jerzy A. Filar, Yuanlie Lin, Lieneke Spanjers. 2004. Stochastic target hitting time and the problem of early retirement. *IEEE Transactions on Automatic Control* **49**(3) 409–419.
- Bouakiz, M., Y. Kebir. 1995. Target-level criterion in markov decision processes. *Journal of Optimization Theory and Applications* **86** 1–15.
- Chiang, W.-C., J.C.H. Chen, X. Xu. 2007. An overview of research on revenue management: Current issues and future research. *International Journal of Revenue Management* **1**(2) 97–128.
- Defourny, Boris, Damien Ernst, Louis Wehenkel. 2008. Risk-aware decision making and dynamic programming. *NIPS 2008 Workshop on Model Uncertainty and Risk in RL*.
- Feng, Y., B. Xiao. 1999. Maximizing revenues of perishable assets with a risk factor. *Operations Research* **47**(2) 337–341.
- Feng, Youyi, Baichun Xiao. 2008. A risk-sensitive model for managing perishable products. *Operations Research* **56**(5) 1305–1311.
- Howard, Ronald A., James E. Matheson. 1972. Risk-sensitive markov decision processes. *Management Science* **18**(7) 356–369.
- Huang, K., K. Chang. 2009. A simulation analysis for the airline seat control considering expected revenue and risk. *Journal of Revenue and Pricing Management* .
- Koenig, Matthias, Joern Meissner. 2008. List pricing versus dynamic pricing: Impact on the revenue risk. Tech. rep., Lancaster University Management School.

- Koenig, Matthias, Joern Meissner. 2009. Risk management policies for dynamic capacity control. Tech. rep., Lancaster University Management School.
- Lai, K.-K., W.-L. Ng. 2005. A stochastic approach to hotel revenue optimization. *Computers and Operations Research* 32(5) 1059–1072.
- Lancaster, J. 2003. The financial risk of airline revenue management. *Journal of Revenue and Pricing Management* 2(2) 158–165.
- Lautenbacher, Conrad J., Shaler Jr. Stidham. 1999. The underlying markov decision process in the single-leg airline yield-management problem. *Transportation Science* 33(2) 136–146.
- Lee, T., M. Hersh. 1993. A model for dynamic airline seat inventory controls with multiple seat bookings. *Transportation Science* 27(3) 252–265.
- Levin, Y., J. McGill, M. Nediak. 2008. Risk in revenue management and dynamic pricing. *Operations Research* 56(2) 326–343.
- Lim, A.E.G., J.G. Shanthikumar. 2007. Relative entropy, exponential utility, and robust dynamic pricing. *Operations Research* 55(2) 198–214.
- Maglaras, C., J. Meissner. 2006. Dynamic pricing strategies for multiproduct revenue management problems. *Manufacturing & Service Operations Management* 8(2) 136–148.
- Markowitz, Harry M. 1952. Portfolio selection. *The Journal of Finance* 7(1) 77–91.
- Mitra, D., Q. Wang. 2005. Stochastic traffic engineering for demand uncertainty and risk-aware network revenue management. *IEEE/ACM Transactions on Networking* 13(2) 221–233.
- Puterman, Martin L. 2005. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley.
- Rockafellar, R. Tyrrell, Stanislav Uryasev. 2000. Optimization of conditional value-at-risk. *Journal of Risk* 2(3).
- Shapiro, Alexander. 2009. On a time consistency concept in risk averse multistage stochastic programming. *Operations Research Letters* 37 143–147.
- Sladky, Karel, Milan Sitar. 2008. Risk-sensitive and mean variance optimality in markov decision processes. *Proceedings of 26th International Conference Mathematical Methods in Economics*.
- Sobel, Matthew J. 1982. The variance of discounted markov decision processes. *Journal of Applied Probability* 19(4) 794–802.
- Talluri, K.T., G.J. van Ryzin. 2005. *The Theory and Practice of Revenue Management*. Springer.

- von Neumann, J., O . Morgenstern. 1947. *Theory of Games and Economic Behavior*. Princeton University Press.
- Weatherford, L.R. 2004. EMSR versus EMSU: Revenue or utility? *Journal of Revenue and Pricing Management* 3(3) 277-284.
- White, D. J. 1974. Dynamic programming and probabilistic constraints. *Operations Research* 22 654-664.
- White, D. J. 1988. Mean, variance, and probabilistic criteria in finite markov decision processes: A review. *Journal of Optimization Theory and Applications* 56(1) 1-29.
- White, D. J. 1993. A survey of applications of markov decision processes. *The Journal of the Operational Research Society* 44 1073-1096.
- Wu, Congbin, Yuanlie Lin. 1999. Minimizing risk models in markov decision processes with policies depending on target values. *Journal of Mathematical Analysis and Applications* 231 47-67.

Appendix

The following theorems are based on the work of Bouakiz and Kebir (1995). The theorems show that our dynamic programming equations find the optimal policy.

Theorem 1. *For each $(c, i) \in S$ and $n \geq 0$, $V_n^{\tilde{\pi}^*}((c, i), \cdot)$ is a distribution function.*

$V_0^{\tilde{\pi}^*}((c, i), \cdot)$ is a distribution function. Assume that $V_{n-1}^{\tilde{\pi}^*}((c, i), \cdot)$ is a distribution function. As by definition

$$V_n^{\tilde{\pi}^*}((c, i), x) = \min_{a \in A} \left\{ \sum_{j=0}^k p_{n-1,j}^r V_{n-1}^{\tilde{\pi}^*}((c-a, j), x - aF_i) \right\}.$$

The sets A, S, R are all finite. The finite convex combination and minimum of distribution functions are distribution functions, too. Thus, $V_n^{\tilde{\pi}^*}((c, i), \cdot)$ is a distribution function. \square

Theorem 2. *For each $n \geq 1$, the sequence $\{V_n^{\tilde{\pi}^*}\}$ satisfies the relations*

$$V_n^{\tilde{\pi}^*}((c, i), x) = \begin{cases} 1, & \text{if } x > \rho_N^{\tilde{\pi}^*}(c, i) \\ \min_{a \in A} \left\{ \sum_{j=0}^k p_{n-1,j}^r V_{n-1}^{\tilde{\pi}^*}((c-a, j), x - aF_i) \right\}, & \text{if } x \leq \rho_N^{\tilde{\pi}^*}(c, i). \end{cases}$$

We define the following operators for convenience. Let a be an arbitrary action and δ_n an arbitrary decision rule at time n . The operators M^a , M^{δ_n} and M are as follows:

$$M^a g_n((c, i), x) = \sum_{j=0}^k p_{n-1,j}^r g_{n-1}((c-a, j), x - aF_i),$$

$$M^{\delta_n} g_n((c, i), x) = M^{\delta_n((c,i),x)} g_n((c, i), x),$$

$$M g_n((c, i), x) = \min_{a \in A} \{M^a g_n((c, i), x)\}.$$

Let $\pi = (\delta_{n-1}, \dots, \delta_1)$ be a policy and δ_n be a decision rule. Let policy $\gamma = (\delta_n, \delta_{n-1}, \dots, \delta_1)$, then $V_n^\gamma((c, i), x) = M^{\delta_n} V_{n-1}^\pi((c, i), x)$ by definition. With $a = \delta_n((c, i), x)$, it is

$$V_n^\gamma((c, i), x) = M^a V_{n-1}^\pi((c, i), x).$$

It follows that

$$V_n^\gamma((c, i), x) \geq M V_{n-1}^\pi((c, i), x) \geq M V_{n-1}^{\tilde{\pi}^*}((c, i), x),$$

and the minimum over y yields

$$V_n^{\tilde{\pi}^*}((c, i), \mathbf{x}) \geq MV_{n-1}^{\tilde{\pi}^*}((c, i), \mathbf{x}).$$

With y , a and $V_n^y((c, i), \mathbf{x}) = M^a V_{n-1}^\pi((c, i), \mathbf{x})$ as before, it is

$$V_n^{\tilde{\pi}^*}((c, i), \mathbf{x}) \leq M^a V_{n-1}^\pi((c, i), \mathbf{x}), \forall \pi.$$

Let $\epsilon > 0$ be arbitrary and λ be a policy which depends on $(c, i), \mathbf{x}, \epsilon$ such that

$$V_{n-1}^\lambda((c, i), \mathbf{x}) \leq V_{n-1}^{\tilde{\pi}^*}((c, i), \mathbf{x}) + \epsilon.$$

As π is arbitrary in the above formula,

$$V_n^{\tilde{\pi}^*}((c, i), \mathbf{x}) \leq M^a V_{n-1}^\lambda((c, i), \mathbf{x}) \leq M^a [V_{n-1}^{\tilde{\pi}^*}((c, i), \mathbf{x}) + \epsilon] \leq M^a V_{n-1}^{\tilde{\pi}^*}((c, i), \mathbf{x}) + \epsilon.$$

Now, since δ_n is an arbitrary decision rule, $a = \delta_n((c, i), \mathbf{x})$, it follows

$$V_n^{\tilde{\pi}^*}((c, i), \mathbf{x}) \leq MV_{n-1}^{\tilde{\pi}^*}((c, i), \mathbf{x}) + \epsilon,$$

and $\epsilon > 0$ is arbitrary,

$$V_n^{\tilde{\pi}^*}((c, i), \mathbf{x}) \leq MV_{n-1}^{\tilde{\pi}^*}((c, i), \mathbf{x}).$$

Hence,

$$V_n^{\tilde{\pi}^*}((c, i), \mathbf{x}) = MV_{n-1}^{\tilde{\pi}^*}((c, i), \mathbf{x}). \quad \square$$

Note, the sequence $\{V_n^{\tilde{\pi}^*}\}$ is monotonically decreasing. It is $V_1^{\tilde{\pi}^*} \leq V_0^{\tilde{\pi}^*}$. Hence, assuming that $V_n^{\tilde{\pi}^*} \leq V_{n-1}^{\tilde{\pi}^*}$, we have $V_{n+1}^{\tilde{\pi}^*} = MV_n^{\tilde{\pi}^*} \leq MV_{n-1}^{\tilde{\pi}^*} = V_n^{\tilde{\pi}^*}$.