Pricing Decisions in Revenue Management: An Experimental Investigation

Ümmühan Akbay
Department of Industrial Engineering, Özyeğin Üniversity, Istanbul 34974, Turkey
Department of Operations Management and Information Systems, Sabancı University, Istanbul 34956, Turkey.
ummuhan.akbay@isikun.edu.tr

Nur Ayvaz Çavdaroğlu
Business Administration Department
Kadir Has University, İstanbul 34083, Turkey
nur.cavdaroglu@khas.edu.tr

Abstract

Behavioral studies in revenue management usually consider only the capacity allocation problem. In this paper, we study both pricing and capacity allocation decisions and report the findings of a decision-making experiment on a two-class revenue management problem. For simplicity we consider a scenario where for the first customer-class the price is predetermined and the demand is (practically) infinite. For the other customer class, the price and the reserved capacity are determined by the decision-makers. Under one treatment, subjects decide on only the price where the reserved capacity is automatically and optimally determined by the software. Under the other treatment, subjects determine both the price and the reserved capacity simultaneously. We consider two different levels for the predetermined first customer class price; high and low. Our findings indicate that on average pricing decisions are not significantly different from the optimal even when the complexity of the task is increased by a second decision. On the other hand, capacity decisions deviate from the optimal significantly and systematically resulting in a "too-low-too-high" pattern.

Keywords
Behavioral Operations, Revenue Management, Pricing and Anchoring

1. Introduction

Revenue management (RM) can be defined as "selling the right product to the right customer at the right time at the right price". The essence of RM lies in price discrimination. Setting the correct prices for several customer segments is especially important in service industries, such as the airline industry, where the profit margins are becoming increasingly narrow. Barnes (2012) argues that "the sale of just one or two more seats on each flight, or a $1 change in the price of each seat sold, can determine profit or loss" of an airline. In such an environment, the accuracy of pricing decisions could be the major factor determining the fate of an entire organization. Moreover, pricing policies have become a fundamental component of manufacturing and service companies' operations in the past two decades due to the increased availability of data, the ease of updating prices via technological enhancements and the availability of decision support tools to enable fast and accurate decisions (Elmaghraby and Keskinocak 2003).

Despite an increasing amount of automation in business processes in almost all industries, it is projected that the human decision-maker will continue to remain as the fine-tuner of the pricing decisions in the foreseeable future (Oancea 2018). Gero Decker (co-founder of Signavio, a Silicon Valley-based software firm) argues that: "AI technology and machine learning technology works great, if you have more than a billion in revenue with your company. If you have less than a billion in revenue for your retailer, the machine learning is simply not powerful enough, and it is much smarter to have people sitting there, building all kinds of rules and decision models for pricing and campaigns and tailoring the price to the specific person." (Mendling et al. 2018) Hence it seems the human intervention and judgement will play a crucial role in revenue management practices of companies in the future as it

does today. Since the pricing policies have an enormous impact on the operations and profitability of all sorts of firms; investigating how human decision makers finalize prices in an aim to attain highest profit margins deserves attention.

Behavioral Operations Management studies human behavior in complex operations contexts through lab and field experiments as well as case studies. The objective of this field is to build a bridge between operations management practice and the theory by testing whether the decisions made by humans follow theoretical expectations, and if not, identifying the factors resulting in this deviation and incorporating these factors into the theory and building new improved models. Although the emergence of the field dates back to 1960s with the famous "beer distribution game" and the "bullwhip effect", the field started to receive increased attention after the seminal paper of Schweitzer and Cachon in 2000. In this paper, the authors demonstrate how decision-makers systematically deviate from the optimal order quantity in a newsvendor setting. Since then, there has been a plethora of behavioral and experimental research in operations management mostly focusing on inventory management and supply chain contracting and in a sense neglecting RM. Nevertheless, there are a few behavioral studies focusing on how human decisions compare to the theoretical models of RM. These studies investigate capacity allocation decisions (Bearden et al. 2008; Bendoly 2011; Kocabıyıkoğlu et al. 2015; Cesaret and Katok 2018). In terms of pricing decisions, the existing behavioral operations literature is restricted to price setting newsvendor model (Kocabıyıkoğlu et al. 2016; Ramachandran et al. 2018) or supply chain contracting (Katok and Wu 2009). Although experimental economics has studied pricing decisions, these studies are limited to the markdown management of the sellers in the presence of strategic customers.

Hence, the pricing models of revenue management, which can be considered one of the biggest successes of operations research, has not received the attention it deserves. To the best of our knowledge, our study is the first attempt to analyze how the decisions of the human decision-makers align with the theoretical findings of the two-class revenue management problem from a price-setting dimension. With this study, we aim to fill the gap in the behavioral operations literature by focusing on the classical, two-class RM problem, whose solution is first characterized by Littlewood (1972). In this classical two-class revenue management problem, the decision maker determines the allocation of capacity between the two classes. This problem has various extensions, e.g. the decision maker could only allocate capacities, or choose the price and protection levels simultaneously. The theoretical problem considered in this study is explained in greater detail in Section 3.

1.1 Objectives

The objectives of this research can be stated as follows:

- To study the classical two-class RM pricing problem where the decision maker needs to set the price of one of the classes while the price of the other class is given, and identify the systematic behavioral patterns in the decisions of the human decision makers;

- To study the classical two-class RM pricing problem with capacity allocations where the decision maker needs to set both the price of one of the classes and the protection level (i.e. allocated capacity) for this class, and identify the systematic behavioral patterns in the decisions of the human decision makers.

Making pricing decisions can be regarded as a more complicated job than merely allocating capacity to customer classes, since it is an art to find the optimal prices that would be appealing to each segment, minimize the cannibalization effects and thereby extract the utmost revenue from the clientele. With this study, we hope to shed light onto the complicated decision-making process of the decision makers in setting prices, and prices and quantity allocations simultaneously. Some of our results are in alignment with the previous literature, while some are new and surprising.

The rest of the paper is organized as follows: in Section 2, we review the relevant literature; in Section 3, we give the details of the theoretical model. Section 4 explains the details of the experimental design and data collection. Section 5 presents the analysis results and Section 6 concludes the paper together with a discussion on the managerial implications of the results.

2. Literature Review

The literature related to behavioral models in revenue management can be started with the work of Bearden et al. (2008), who consider a decision maker's problem of selling a number of identical products under uncertain demand. In their setting, the customers vary in their willingness-to-pay for the product and each arriving customer makes a

price offer to purchase one unit of the product. The seller needs to decide whether to accept or reject each offer. The authors designate three sets of experiments that differ in the problem parameters (e.g. the arrival rate), and the results emphasize the bounded rationality of the decision makers under complex environments. Bendoly (2011) uses a similar set-up to examine the effect of feedback on the responses of decision makers engaged in RM tasks by measuring their physiological reactions. The author argues that the framing of the feedback is significant in determining the decisions of the subjects and the obtained revenue levels.

Another dynamic revenue management setting is studied by Cesaret and Katok (2018). They consider classical two-class problem of revenue management. They designate a behavioral setting where the human decision makers must decide whether to accept or reject each arriving customer belonging to either of the two classes. The authors consider various experiment settings with or without ordered arrivals of the customers belonging to each class, and with or without making the accept decision upfront. Analyzing the decision-making patterns of the participants in each setting, they develop a behavioral choice model incorporating the regret element. However, their study differs from ours mainly because they focus on a problem with fixed prices and investigate the behavior in capacity allocation RM decisions

The rest of the studies are based on a static setting similar to ours. Kocabıyıkoğlu et al. (2015) study the two-class RM problem and a mathematically equivalent newsvendor problem, and test whether the behavior in these two settings would be the same in practice. They find that revenue management allocation decisions are consistently higher than the newsvendor order quantities. The authors argue that the reason for this behavior is the nature of the overage cost in the two settings, emphasizing the importance of the framing effect. In another paper, the same authors present an experimental study of the price-setting newsvendor problem, which allows the decision maker to determine both the selling price and the order quantity (Kocabıyıkoğlu et al. 2016). They find evidences of anchoring behavior on the expected demand when quantity is the decision variable and on the initial inventory level when price is the decision variable. Kocabıyıkoğlu et al. (2018) investigate the decision-making process in classical two-class RM problem and find that subjects' behavior is influenced by the type of the decision (booking limit versus protection level) they need to make. In particular, the number of units reserved for the high-end segment vary across the decision types. Although they consider the same problem as we do, the pricing decision is not involved in this study.

Ramachandran et al. (2018) experimentally assess the subjects' performance when they are making joint quantity and price decisions, which is prevalent in fashion, medicine and publishing industries. The deviations of the decision makers from the optimal price and quantity levels are explained by models that involve reference-dependence and anchoring effects. These effects are alleviated when the uncertainty is removed or information about the market demand is provided to subjects after they decide the price level. Their study is similar to ours in the sense that they also consider a joint decisions setting. However, their study investigates the differences between simultaneous and sequential joint decisions rather than one of the decisions being automatically computed as in our setting.

As evident from the above review, the behavioral studies in revenue management mostly focus on the capacity allocation decision. Pricing decisions, if involved, are only considered together with quantity decisions. Regarding the experimental studies in pricing, one should shift attention to the economics literature. However, most of the studies in this literature stream focus on the behavior of the customers or the markdown management strategies of the firms in the presence of strategic buyers. Reynolds (2000) presents a laboratory study of dynamic pricing in a durable-goods monopoly with a single human seller and 1–5 human buyers and observes that some buyers postpone their purchases in accordance with a strategic behavior. Moreover, the sellers consistently set higher prices in that case. Cason and Mago (2013) designate an experimental setting in which two sellers compete to sell to a single buyer, and find that while the sellers underprice in a single-round game, they often overprice in the multiple-rounds game. Mak et al. (2012) consider a pricing experiment in which the offers are made by the sellers sequentially to sell a single unit of the good to the same buyer. They find that prices are generally consistent with the model predictions if the buyer is myopic, but significantly above optimal if the buyer is strategic. Finally, Kremer et al. (2017) both analytically develop and experimentally test a two-period pricing model in the presence of strategic customers, and observe that sellers underprice when strategic customers comprise a considerable portion of the customer base.

Our problem is also related to the studies investigating the anchoring and reference-dependence effects in operational management settings. Kahneman and Tversky (1979) define anchoring effect as "disproportionate influence on decision makers to make judgments that are biased toward an initially presented value". That is, people make insufficient adjustments to produce a final estimation, if they are presented with an initial value of the same parameter.

Moreover, several studies have shown that the anchoring effect is not reduced significantly by an increasing expertise in the field (Mussweiler et al. 2000; Englich et al. 2006), and the individual differences might affect the rate of the anchoring bias on the decision-making process.

The quantity decisions in newsvendor settings are often shown to be biased by the demand chasing heuristic, which is defined as a systematic adjustment of the order quantities towards the previous period's demand values (see e.g. Ben-Zion et al. 2008, Bostian et al. 2008, Lurie and Swaminathan 2009, Schweitzer and Cachon 2000). We also explore the effect of this bias factor in the decisions of the subjects in our experiments directly in the protection level decisions and indirectly in the pricing decisions through the induced protection levels.

3. Theoretical Model

We consider the typical two-class revenue management problem (Littlewood 1972; Belobaba 1989) where the decision maker determines the allocation of capacity between the two classes. In other words, the decision-maker determines how much capacity to reserve for the higher priced class, which is called "the protection level".

We assume that the demand for the class 1 is ample enough to be assumed practically infinite whereas the demand for class 2 is stochastic and dependent on the class 2 price. Class 1 price is fixed leaving only class 2 price and protection level for the decision maker to determine. Parameters of the model are defined as follows:

C: Total capacity to be allocated between the two customer classes

p_i: Class i price, $i=1,2$ where $p_2 \geq p_1$

D_i: Demand for class i, $i=1,2$

We assume $D_1 = \infty$ and D_2 distributed with $F_2(\cdot\,|p_2)$ where,

$F_2(\cdot\,|p_2)$: The cumulative distribution function for class 2 demand given price p_2

Under these assumptions the problem of the decision-maker is to find the optimal class 2 price p_2 that will maximize the below expected revenue:

$$\pi(p_1, p_2) = p_2 E\left[\min\{D_2, PL^*(p_1, p_2)\}\right] + p_1(C - PL^*(p_1, p_2))$$

Here $PL^*(p_1, p_2)$ is the optimal protection level for class 2 given class prices p_1 and p_2. This problem is solved in two steps; in the first step p_1 and p_2 are assumed fixed and optimal protection level $PL^*(p_1, p_2)$ is derived. This is equivalent to finding the optimal order quantity in the newsvendor problem which leads to the following result:

$$PL^*(p_1, p_2) = F_2^{-1}\left(1 - \frac{p_1}{p_2}\middle|p_2\right)$$ (Littlewood 1972).

In the second step, the derived $PL^*(p_1, p_2)$ is inserted in the expected revenue formulation and optimal class 2 price p_2 is derived.

This problem is a simplified version of the pricing problem used in several real-life instances where the decision-maker needs to determine the prices for a number of products simultaneously. In the above problem, the price for class 1 (or the low-end segment) is fixed, and the pricing decision is not dynamic, which makes it a much simpler version of the real-life problem. However, even in this simplified case, whether the decision makers manage to set the correct price at all periods remains to be a wonder, and we hope to explore the systematic errors that the decision makers make in an attempt to set the optimal price.

4. Experiment Design and Data Collection

We base our experiment scenario on a two-class airline revenue management context since airline industry is the major user of the revenue management practices. Specifically, we ask the subjects to maximize the total revenue by determining the price and/or the protection level for the high-end segment given that the low-end segment price is fixed and known.We employ a 2 × 2 between-subjects design. The treatments are based on decision type (single vs joint decisions) and class 1 price level (low vs high). In the Single Decisions (SD) treatments, subjects are asked to determine class 2 price while the protection level set optimally and automatically by the software. On the other hand, subjects who participate in the Joint Decisions (JD) treatments determine both class 2 price and the protection level simultaneously.

In all treatments, the capacity of the airplane which is to be allocated between class 1 and class 2 is fixed as 200 seats. The demand of class 1 is assumed to be infinite leading to all of the seats left to class 1 being sold for sure. On the other hand, the demand of class 2 depends on class 2 price. Specifically, given class 2 price p_2 which is determined by the subjects, the demand for class 2 is given by the function $D_2 = 250 - 2p_2 + \epsilon$ where ϵ is assumed to be discrete uniform random variable on the interval $[0,100]$, effectively making class 2 demand a discrete uniform random variable on the interval $[250 - 2p_2, 350 - 2p_2]$. This demand function restricts class 2 price to be at most 125. The optimal protection level is derived as:

$$PL^*(p_1, p_2) = 250 - 2p_2 + 100\left(1 - \frac{p_1}{p_2}\right)$$

Here $\left(\mathbf{1} - \frac{\boldsymbol{p_1}}{\boldsymbol{p_2}}\right)$ determines the profit margin of class 2. When class 1 price is low, the profit margin is high and vice versa. Class 1 price is set to be $\boldsymbol{p_1 = 30}$ for *High Margin* condition and to $\boldsymbol{p_1 = 90}$ for *Low Margin* condition.

Naturally, these prices constitute the lower bound for class 2 prices. The optimal class 2 prices computed according to (1), the corresponding protection levels and expected revenue values are summarized in Table 1.

Table 1. Theoretical benchmarks

	High Margin	Low Margin
Class 2 price decision interval	[30, 125]	[90, 125]
Optimal class 2 price	89	112
Optimal protection level	138	45
Optimal expected total revenue	12252	18800

The parameters used in the experimental design are selected based on actual information and parameter settings in the literature. The most common commercial airplane models Airbus A320 and Boeing 737 have a capacity to hold maximum 180 or 189 seats. With the aim to make mental accounting easier for the experiment subjects, these seat capacities are rounded to 200 in our scenario. The demand function and class 1 prices are chosen from parameter setting used in Kocabiyikoglu et al. (2015) and adjusted according to our setting.

Experiments are conducted at an emerging research university following standard ethical guidelines for experiments with human subjects. A total of 169 undergraduate students from the student body volunteered to participate in the study. Out of the 169 subjects, 30% are male and 70% are female; 39% are from Engineering, 56% are from Social Sciences and the rest is from various departments including Management. We checked the effect of subject's education background gender and found no significant difference. Each subject participated in exactly one treatment. A summary of the experimental conditions and the number of subjects in each treatment are given in Table 2.

Table 2. Sample sizes of the treatments

	Single Decision	Joint Decisions
High Margin	42	44
Low Margin	41	42

Subjects are recruited through an online recruitment system where they can earn a fixed credit for the course of their selection from the list of credit offering courses. The subjects of this study are granted 1% course credit and no additional incentive is provided.

The same protocol is followed for all sessions. Before the experiment, written instructions of the experiment scenario are sent to the subjects and they are asked to come to the experiment having read the instructions. Right before the experiment, a brief tutorial about the experiment scenario is given by the experimenter and the experiment software is explained. After the subjects' questions are answered, the experiment is started. Each session lasted about 50 minutes including the tutorial. During the experiment, subjects are not allowed to communicate with each other.

The experiments are conducted using Visual Basic for Applications (VBA) with MS Excel interface. Information regarding the total capacity, class 1 price and the range of the class 2 price is placed on the upper left part of the screen. Subject enter their decisions in the box on the upper right corner. When the subjects enter the class 2 price, the range for the class 2 demand is computed on the screen. Then the subjects determine the protection level from this demand range and do a "what-if" analysis using the decision support tool provided in the middle of the screen. This decision

support tool reports sold and unsold seats, and the revenue values for a set of equidistanced values of the demand range so that the subjects can see the best and worst cases and get a feeling of the expected revenues. After the what-if analysis, the pricing and protection level decisions are submitted through the "Submit Decisions" button. The demand realization, sold seats, unsold seats, as well as total revenue of the current period and cumulative revenue are computed and displayed on the table at the lower part of the screen. The information regarding all past data is available to the subjects at all times. The subjects made decisions for 43 consecutive periods. First three periods are reserved for warm-up, which is not included in the analysis, to help the subjects get acquainted with the interface and the task.

5. Results and Discussion

5.1 Comparison with Theoretical Benchmarks

Here we present the comparison of the experiment results with the theoretical benchmarks. Experiment performance of each subject is averaged over 40 periods. Hence, each subject constitutes one data point and the sample sizes for the hypothesis tests are the number of subjects in each treatment. We use two tailed Wilcoxon signed rank test (Siegel, 1956) for the comparisons.

Single decisions treatment

Hypothesis 1: *In the Single Decisions (SD) treatment for both profit margins, subjects' pricing decisions will be as predicted by theory.*

Table 3 compares experiment performances with the theoretical benchmarks for the SD treatment. The pricing decisions of the subjects in both Low and High Margin conditions are found to be not significantly different from the optimal price values. Consequently, the automatically computed protection levels induced by subjects' pricing decisions are not significantly different from the optimal protection levels. However, the difference between the theoretical prediction and the average expected and realized revenues, though small (3.4% and 0.5%), are significant. Individual deviations, deviations in each period and the nonlinear nature of the revenue function may be the reasons why the experiment revenues fall below the optimal.

Table 3. Comparison results of experiment performance with the optimal thresholds (SD)

		Experiment Data			Optimal	P Value
		Mean	Std. Dev.	Median		
High Margin	Class 2 Price	88.46	9.58	90.60	89	0.78
	Protection Level	137.41	14.83	134.98	138	0.78
	Expected Revenue	11674.66	469.00	11833.85	12251.85	**0.00**
	Realized Revenue	11697.49	505.05	11830.21	12251.85	**0.00**
Low Margin	Class 2 Price	111.50	6.73	113.68	112	0.25
	Protection Level	45.68	7.95	43.05	45	0.34
	Expected Revenue	18648.30	133.51	18702.75	18800.38	**0.00**
	Realized Revenue	18648.44	146.87	18692.80	18800.38	**0.00**
P-values are obtained from two tailed Wilcoxon signed rank test. Sample size is the number of subjects.						

A subject level analysis presented in Table 4 reveals that more than 2/3 of the subjects' pricing decisions are in fact significantly different from the optimal. Additionally, period averages of class 2 prices and the induced protection levels visualized in Figure 1 show that there is significant difference between subject decisions and the optimal even on the average level. These differences result in the deviations in the expected revenue values presented in Figure 2. Furthermore, Figure 1 shows that at the start of the experiment, subjects' pricing decisions are anchored at class 1 price, and as the experiment progresses, the prices are adjusted towards the optimal indicating presence of an "anchoring and adjustments" type decision at the beginning of the experiment.

Table 4. Individual level comparison of pricing decisions with the optimal (SD)

	# Subjects with P-value<0.05	# Subjects with P-value>0.05
High Margin	29	13
Low Margin	28	13

Figure 1: Average class 2 price and protection level over the periods (SD)

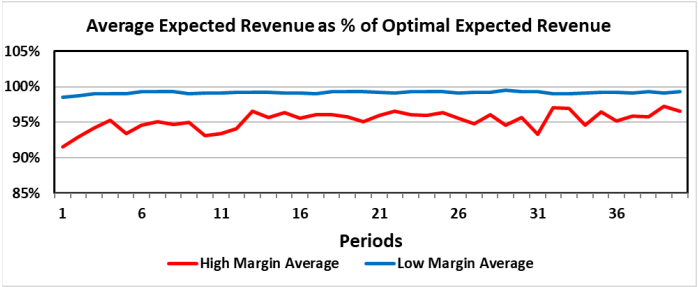

Figure 2: Average expected revenue as percentage of the optimal over the periods (SD)

Overall, Hypothesis 1 finds partial support on the average data; however, individual level analysis indicates that the hypothesis is to be rejected.

Joint decisions treatment

Hypothesis 2a: *In the Joint Decisions (JD) treatment for both profit margins, subjects' pricing decisions will be as predicted by theory.*
Hypothesis 2b: *In the Joint Decisions (JD) treatment for both profit margins, subjects' protection level decisions will be as predicted by theory.*

Comparison results presented in Table 5 show that the pricing decisions of the subjects are again not significantly different from the theoretical optimal. (Although, for the High Margin condition, the difference is almost significant as the p-value is 0.052.) On the other hand, the protection level decisions seem to be significantly below the optimal for the High Margin condition and significantly above the optimal for the Low Margin condition. As a result, expected and realized revenue values deviate from the theoretical prediction significantly.

Table 5. Comparison of joint decisions treatment results with the optimal (JD)

		Experiment Data			Optimal	P-Value
		Mean	Std. Dev.	Median		
High Margin	Business Class Price	92.57	10.44	92.24	89	0.05
	Protection Level	118.86	22.79	118.24	138	**0.00**
	Expected Revenue	11283.24	491.38	11323.52	12251.85	**0.00**
	Realized Revenue	11352.57	486.17	11394.86	12251.85	**0.00**
	Business Class Price	110.66	5.99	111.36	112	0.18

Low Margin	Protection Level	77.75	19.09	75.64	45	**0.00**
	Expected Revenue	17752.65	670.67	17913.78	18800.38	**0.00**
	Realized Revenue	17801.47	661.37	17969.50	18800.38	**0.00**
P-values are obtained from two tailed Wilcoxon signed rank test. Sample size is the number of subjects.						

Further analysis of individual decisions, of which the results are presented in Table 6 show that for more than about 3/4 of the subjects, the pricing and protection level decisions are significantly different from the optimal. (Protection level decisions are compared with the optimal protection level corresponding to the pricing decision of the subjects in each period.)

Table 6: Individual level comparison of pricing and protection level decisions with the optimal (JD)

		# Subjects with P-value<0.05	**# Subjects with P-value>0.05**
Pricing Decisions	Low Economy Class Price	36	8
	High Economy Class Price	31	11
Protection Level Decisions	Low Economy Class Price	35	9
	High Economy Class Price	41	1
P-values are matched sample Wilcoxon signed rank test			

Figure 3 reveals that evidence of anchoring at class 1 price is still present in the Low Margin condition; however, both the anchor and the adjustments seem weaker than that of the SD treatment. For the High Margin condition, prices do not seem to be anchored at class 1 price at all, and yet the tendency to increase the price over time can still be observed.

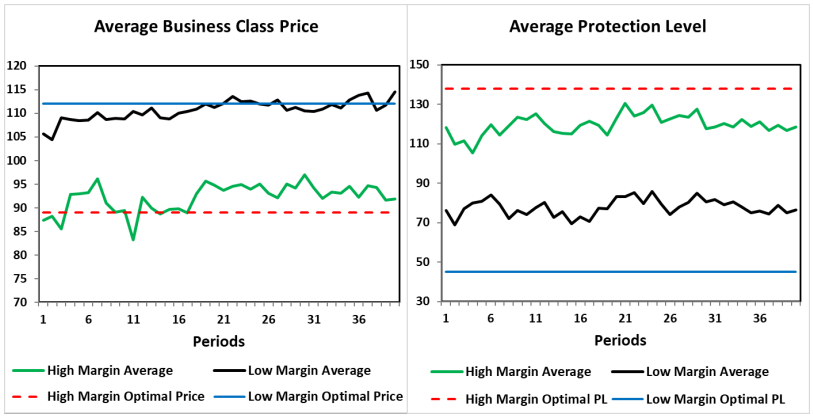

Figure 3: Average class 2 price and protection level over the periods (JD)

The comparison results presented in Table 5 along with the period average of the decisions displayed Figure 3 reveal an interesting "too low-too high" pattern. Specifically, subjects' protection level decisions are systematically below the optimal (too low) in the High Margin condition and above the optimal (too high) in the Low Margin condition. To paraphrase, our subjects underprotect when profit margin is high and overprotect when profit margin is low. This behavior resembles and parallels the "too low-too high" pattern observed in newsvendor experiments (Schweitzer and Cachon 2000; Bolton and Katok 2005; Kocabıyıkoğlu et al. 2016; Surti et al. 2020). In the newsvendor context, this pattern corresponds to subjects consistently ordering quantities lower than the optimal for high-profit products and higher than the optimal for low-profit products. This pattern is due to the subjects' tendency to "anchor" around demand average in the early periods of the experiment and adjust insufficiently in the subsequent periods towards the optimal order quantity. To check for such tendency in our data, we plot average protection levels along with the average demand mean and average optimal protection level corresponding to subjects' price decisions $PL^*(p_2)$ in Figure 4. For High Margin condition, average protection level decisions seem to start below the demand mean, then

catch up with it around the 5th period, and halfway through the experiment adjust towards the optimal. For Low Margin condition, the anchor around the demand mean is more salient and persists throughout the experiment.

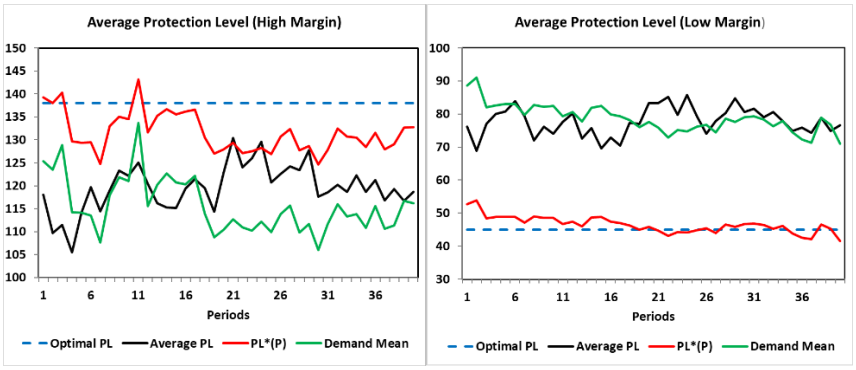

Figure 4: Average protection level, demand mean and optimal protection level

To conclude, similar to Hypothesis 1, Hypothesis 2a finds partial support on the average data, but is rejected by the individual level analysis. Hypothesis 2b, on the other hand, is rejected both on the average and individual levels.

5.2 Comparison of Single and Joint Decisions

Here, we compare the experiment results of the single and joint decisions treatments. The analysis is conducted using two-tailed Mann Whitney U test.

Hypothesis 3a: *The difference between the pricing decisions of the single and joint decisions treatments will not be significant.*

Hypothesis 3b: *The difference between the protection level decisions of the single and joint decisions treatments will not be significant.*

The results presented in Table 7 show that all but one differences between the SD and JD treatments are significant. Under High Margin condition, JD treatment results in higher class 2 prices, lower protection levels and lower revenues. Under Low Margin condition, JD treatment leads to lower (on average) prices, higher protection levels and lower revenues.

Table 7. Comparison of single and joint decision treatments

		Single Decisions Median	Joint Decisions Median	P-Value
High Margin	Class 2 Price	90.60	92.24	**0.04**
	Protection Level	134.98	118.24	**0.00**
	Expected Revenue	11833.85	11323.52	**0.00**
	Realized Revenue	11830.21	11394.86	**0.00**
Low Margin	Class 2 Price	113.68	111.36	0.13
	Protection Level	43.05	75.64	**0.00**
	Expected Revenue	18702.75	17913.78	**0.00**
	Realized Revenue	18692.80	17969.50	**0.00**

P-values are obtained from one tailed Mann Whitney U test. Sample size is the number of subjects.

At first glance, all performance measures seem to become worse off when the subjects are required to make simultaneous decisions rather than a single one. However, going back to Figure 2, we observe higher variability in subjects' decisions at the beginning of the experiment. Hence, a comparison based on the second half of the experiment might lead to a better verdict on the issue. In this respect, when subject averages of the class 2 prices over the last 20 periods are compared, the difference between SD and JD treatments is not significant. (P-value is 0.12 for High Margin condition, and 0.28 for Low Margin condition.) This can be interpreted as subjects' (pricing) decisions are not necessarily worsening when subjects are to make two decisions simultaneously instead of one, despite the increasing complexity of the decision problem. Kocabıyıkoğlu et al. (2016) also find similar results. When they compare pricing decisions under pricing-only treatment and price-and-quantity treatments, they find no significant difference. Another observation is that subjects appear to be better at pricing decisions than protection level decisions. The difference

between average pricing decision and the optimal is much smaller compared to the gap between the protection level and its corresponding optimal. This is also supported by the findings of Kocabıyıkoğlu et al. (2016).

Overall, we fail to reject Hypothesis 3a, while Hypothesis 3b is strongly rejected.

5.3 Discussion

In the light of the above findings, the main results and insight obtained from our experiments are as follows.

• The decision makers are better at the pricing decisions compared to setting the protection levels. No significant difference was found in the pricing decisions between single decisions treatments and joint decisions treatments after the initial "learning phase" is over. Despite the fact that the expected and realized revenue values deteriorate in joint decisions treatments, this result is not due to bad decision making in pricing, but due to the fact that the decision makers could not set the protection levels correctly. Hence, although making multiple decisions at once seems to decrease the performance of decision-makers, the quality of all types of decisions need not be worsened.

• The deviations in the protection level decisions exhibit a too low/too high pattern. Namely, the subjects underprotect for the high-end customer class when the profit margin is high, and overprotect when the profit margin is low. A possible explanation for this observation is anchoring towards the mean demand in both cases. A similar phenomenon is observed in newsvendor settings when decision makers set the ordering quantity. On the other hand, demand chasing heuristic or gambler's fallacy are not found to be much significant in explaining the decision patterns of the subjects. That is, rather than previous period's demand values, the decision makers are more affected by the mean demand value in setting the protection levels.

• The impact of the learning is not very clear in our experiments. Although the subjects managed to improve some decisions over time, some other decisions were worsened with experience. Hence, it can be argued that the learning effect does not always positively affect the performance in complex decision-making settings. In accordance with literature, the subjects might become overconfident with their incorrect decisions over time, which leads to a systematic underperformance (Ren and Croson 2013).

The overall findings summarized above lead to the conclusion that the decision makers are not very good at making simultaneous decisions, especially capacity allocation decisions, in complex environments. Hence, as argued in earlier studies, decision support tools will be helpful in improving the performance of the decision makers in such environments (Ramachandran et al. 2018). However, rather than providing the decision makers with classic decision support systems that only provide demand forecasts and other relevant information, it might be more beneficial to induce tools that also detect consistent deviations in one's decisions over time and inform the decision maker accordingly. These deviations are often in the form of an anchoring behavior, i.e. being affected by the irrelevant values present in the system (e.g. the price of a similar product, the average demand of the past periods, etc.) Moreover, the companies could establish training programs for the decision makers to make them aware of the possible behavioral biases as the ones noted in this study. In some complex environments, even automatizing some of the decisions might be considered to improve the overall decision quality. Hence, our results indicate that using smarter decision support tools which show the consistent deviations in one's decisions or even make some decisions automatically could be beneficial in complex, joint decision-making settings.

6. Conclusion

In this article, we present an experimental study of the pricing strategies in classical two-class revenue management problem. Our study contributes to the behavioral revenue management literature by investigating the alignment of pricing decisions of human decision makers with the theoretical RM benchmarks.

The main takeaway from our study is that the number of decisions to be made at each instance affects the decision-making process. In particular, although our subjects manage to make pricing decisions similar to the values prescribed by the normative theory, they performed poorly in setting the protection levels when they had to make joint decisions.

It is true that advanced optimization techniques and software are used in pricing and revenue management and the amount of automation is increasing day by day. However, human judgement is still a crucial component of the decision process in all business contexts. Even in big airline companies who have access to highly technological tools, the

findings of the airline revenue management software are not applied without human supervision. Human decision-makers interfere with the software and change parameters depending on the season or special events (Oancea 2018). Additionally, it is not possible to include all local and special factors into the software and these adjustments require human judgement. Furthermore, many small companies such as hotels or restaurants do not have access to expensive optimization software and most of their pricing decisions are made by humans (Mendling et al. 2018). Thus, our results can be valuable in determining the form of aid for human decision makers in their pricing and capacity allocation decisions in aforementioned contexts.

The results presented in this article can be extended in a number of ways. A valuable extension of our work might be to conduct a similar study with more than two-classes. In those models, the pricing decisions could get complicated quickly and several behavioral biases might be observed in the pricing decisions of the real decision makers. Another interesting extension could be a dynamic pricing experiment where the subjects need to keep track of the demand distribution over the selling period. All the past work in this topic (e.g. Cesaret and Katok, 2018; Bearden et al. 2008) focus on the sell or not sell decision of the decision maker. Thus, it might be interesting to compare their findings with the results of a study where the decision maker has to define the selling price in each period and observe the demand accordingly. It might also be valuable to study how the rate of variation in other problem parameters such as demand and capacity could affect the decisions of the decision makers. We hope that this study attracts further attention to the behavioral analysis of pricing problems of the revenue management.

Acknowledgements

We gratefully acknowledge the support of the Kadir Has University Individual Research Fund (Project ID 2017-BAP-08). We thank Ms. Ece Dal for her valuable contribution in conducting the experiments and in data analysis.

References

Barnes, B., Airline pricing, *The Oxford Handbook of Pricing Management*. London: Oxford University Press. 2012.

Bearden, J. N., Murphy, R. O., and Rapoport, A., Decision Biases in Revenue Management: Some Behavioral Evidence, *Manufacturing and Service Operations Management*, vol 10, pp. 625-636, 2008.

Belobaba, P. P., OR practice—application of a probabilistic decision model to airline seat inventory control, *Operations Research*, vol. 37, pp. 183-197, 1989.

Bendoly, E., Real-time feedback and booking behavior in the hospitality industry: Moderating the balance between imperfect judgment and imperfect prescription, *Journal of Operations Management*, vol. 31, pp. 62-71, 2013.

Cason, T. N., and Mago, S. D., A laboratory study of duopoly price competition with patient buyers, *Economic Inquiry*, vol. 51, pp. 1123-1141. 2013.

Cesaret, B., and Katok, E., A behavioral study of capacity allocation in revenue management, *Working Paper*, The University of Texas at Dallas, 2018.

Elmaghraby, W., and Keskinocak, P., Dynamic pricing in the presence of inventory considerations: Research overview, current practices, and future directions, *Management Science*, vol. 49, pp. 1287-1309, 2003.

Englich, B., Mussweiler, T., and Strack, F., Playing dice with criminal sentences: The influence of irrelevant anchors on experts' judicial decision making, *Personality and Social Psychology Bulletin*, vol. 32, pp. 188-200, 2006.

Kahneman, D., and Tversky, A., Prospect Theory: An analysis of decision under risk, *Econometrica*, vol. 47, pp. 263-291, 1979.

Katok, E., Wu, D. Y., Contracting in Supply Chains: A Laboratory Investigation, Management Science, 55, 1953-1968, 2009.

Kocabıyıkoğlu, A., Göğüş, C. I., and Gönül, M. S., Revenue management vs. newsvendor decisions: Does behavioral response mirror normative equivalence? *Production and Operations Management*, vol. 24, pp. 750-761, 2015.

Kocabıyıkoğlu, A., Göğüş, C. I., and Gönül, M. S., Decision making and the price setting newsvendor: Experimental evidence, *Decision Sciences*, vol. 47, 157-186, 2016.

Kocabıyıkoğlu, A., Göğüş, C. I., and Hekimoğlu, M. H., The Impact of Decision Types on Revenue Management Decisions: An Experimental Study, *Decision Sciences*, vol. 49, pp. 225-249, 2018.

Kremer, M., Mantin, B., & Ovchinnikov, A., Dynamic pricing in the presence of myopic and strategic consumers: Theory and experiment, *Production and Operations Management*, vol. 26, pp. 116-133, 2017.

Littlewood, K., Forecasting and control of passenger bookings, *AGIFORS*, 12, pp. 95-117, 1972.

Mak, V., Rapoport, A., and Gisches, E. J., Competitive dynamic pricing with alternating offers: Theory and experiment, *Games and Economic Behavior*, vol. 75, pp. 250-264, 2012.

Mussweiler, T., Strack, F., and Pfeiffer, T., Overcoming the inevitable anchoring effect: Considering the opposite compensates for selective accessibility, *Personality and Social Psychology Bulletin*, vol. 26, pp. 1142-1150, 2000.

Oancea, O., and Horga, R., The new wave of airline revenue and pricing management–what next for the leisure customer?, *Journal of Revenue and Pricing Management*, vol. 17, pp. 182-188, 2018.

Ramachandran, K., Tereyağoğlu, N., and Xia, Y, Multi-dimensional decision making in operations: An experimental investigation of joint pricing and quantity decisions, *Management Science*, vol. 64, pp. 5544-5558, 2018.

Ren, Y., and Croson, R., Overconfidence in Newsvendor Orders: An Experimental Study, *Management Science*, vol. 59, pp. 2502-2517, 2013.

Reynolds, S. S., Durable-goods monopoly: laboratory market and bargaining experiments, *The RAND Journal of Economics*, vol. 31, pp. 375-394, 2000.

Schweitzer, M. E., and Cachon, G. P., Decision bias in the newsvendor problem with a known demand distribution: Experimental evidence, *Management Science*, vol. 46, pp. 404-420, 2000.

Surti, C., Celani, A., and Gajpal, Y., The Newsvendor Problem: The Role of Prospect Theory and Feedback, *European Journal of Operational Research*, vol. 287, pp. 251-261, 2020.

Biographies

Ümmühan Akbay is an Assistant Professor at Management department of Işık University, İstanbul, Turkey. She received her Ph.D. in Industrial Engineering from Sabancı University. She holds an M.Phil degree in Operations Research from Columbia University Graduate School of Business and a B.Sc. degree in Industrial Engineering from Bilkent University. She was a visiting professor at the Industrial Engineering program, Özyeğin University between 2017-2020. Her research interests include behavioral operations management, behavioral and experimental economics, supply chain management, game theory, decision analysis, energy markets, healthcare operations management and revenue management.

Nur Ayvaz-Çavdaroğlu is an Assistant Professor of Management at Kadir Has University. She received her BS degree in Industrial Engineering from Middle East Technical University and her MSc and Ph.D degrees from the Department of Industrial Engineering & Operational Research of the Columbia University. She worked as a post-doc researcher at Whitman School of Business at Syracuse University. Her research focuses on applications of revenue management and pricing in health care, hospitality and agriculture sectors. She is also interested in supply chain management problems. Her work has appeared in journals such as Production and Operations Management, Omega, Journal of Revenue and Pricing Management, Journal of Cleaner Production, Journal of Travel Research and also as book chapters.