Innovative Applications of O.R.

Joint demand and capacity management in a restaurant system

Johye Hwang\textsuperscript{a}, Long Gao\textsuperscript{b}, Wooseung Jang\textsuperscript{c,}\textsuperscript{*}

\textsuperscript{a}College of Hotel and Tourism Management, Kyung Hee University, Seoul, South Korea
\textsuperscript{b}Anderson Graduate School of Management, University of California, Riverside, CA 92521, USA
\textsuperscript{c}Department of Industrial and Manufacturing Systems Engineering, University of Missouri-Columbia, Columbia, MO 65211, USA

\textbf{A R T I C L E  I N F O}

Article history:
Received 9 March 2009
Accepted 1 April 2010
Available online 7 April 2010

Keywords:
Capacity management
Pricing
Restaurant system
Competitive strategy
Queueing theory

\textbf{A B S T R A C T}

This paper studies the demand and capacity management problem in a restaurant system. A queueing-based optimization model with underlying quasi birth-and-death process and state-dependent functions is developed to address the dynamic and nonlinearity difficulties. With this framework, we empirically examine the relative performance of commonly used strategies for the case of a local restaurant. The study shows that a strategy that balances service quality and cost yields maximum profit. The result indicates that the traditional view of the conflict between service quality and cost can be overcome by using an interdisciplinary perspective of marketing and operations. Both perspectives should be embraced in academic research and industrial practice in capacity planning decisions for services.

\textcopyright 2010 Elsevier B.V. All rights reserved.

\section{1. Introduction}

In service organizations, capacity decisions are made from various strategic perspectives. From a marketing perspective, organizations often set a capacity level that can ensure high quality service. The service level is usually specified as the average steady state waiting time or as a given (e.g. 95\%) percentile of the waiting time distribution. Many restaurant chains and fast food industry outlets use waiting time standards as an explicitly advertised competitive edge. For example, Dominos offers free pizza if the delivery is more than 30 minutes after placing the order. Black Angus provides free lunch if the waiting time is more than 10 minutes. On the other hand, from an operations perspective, service organizations, such as fast food restaurants and health care providers, tend to choose a capacity level that minimizes system costs without decreasing system revenue and quality (Davis, 1991; Kurt and Cote, 1998). Fast food restaurants, for example, carry a high labor cost relative to capital cost and their labor costs affect their profitability. Thus, companies, such as Taco Bell, are concerned about minimizing labor cost in their determination of the optimal allocation of labor (Hueter and Swart, 1998). However, for most of profit driven firms, an integrated perspective that combines both marketing and operations should be taken in order to balance the costs of enhancing services and the increased revenue from improved service level.

Different pricing-capacity policies arise, depending on the competitive strategy and the service level firms choose. If the additional capacity reduces the waiting time, adding capacity may attract additional customers during busy periods and would derive additional revenue. On the other hand, a lack of capacity, despite lower cost, may lead to customers waiting longer or to serving fewer customers. For example, in a situation in which a customer enters a restaurant and sees a long waiting line, the customer may leave the restaurant and look for a less busy one. What makes matters worse is that this customer may not choose the restaurant for another visit. In this paper, we develop a queueing-based optimization model to address a demand and capacity management problem in a service system with service and price-sensitive random demand, and examine the implications of the competitive strategies on its pricing-capacity decisions.

Besides attributes, such as location, ambiance and the quality of food (Auty, 1992; Jaksa, 1997), the price and the waiting time are the most important factors for a customer in selecting a restaurant (Li and Lee, 1994). However, most of the prior models in the marketing and operations literature assume that price and waiting time can be aggregated into a so-called full price, a single variable usually defined as the direct price plus a multiple of the expected waiting time (for example, see Cachon and Harker, 2002 and Chen and Wan, 2003). This implies that (1) all the customers experience the same amount of waiting time, regardless of the current congestion of the restaurant, and that (2) the cost of waiting is proportional to the waiting time.

Research in marketing, operations, economics, and psychology has shown that both of these assumptions are often violated. This
is particularly true in the restaurant industry. For example, at a local Olive Garden restaurant, the waiting time for a customer on a Friday night can vary from zero to two hours, with the restaurant still keeping the service level at an overall target level. However, each customer makes the joining decision based on his own actual waiting, not on the average waiting time. Such heterogeneity in waiting time plays a dominant role in each customer's joining decision, and ultimately determines the demand volume for the restaurant. For the assumption of linear waiting cost, Larson (1987) argued that customer “disutility” of waiting is highly nonlinear in many service industries. Carmon et al. (1995) also highlighted such nonlinearity. Therefore, in staff scheduling decisions, the dynamic congestion (real-time, rather than average) and the nonlinearity of waiting should be explicitly taken into account. In our model, we address these two problems via an underlying Markov process along with a state (congestion) dependent demand rate function and a profit (cost) rate function, where the state represents the number of customers in a restaurant.

To address the questions of staff scheduling and nonlinearity of waiting times, we first develop an optimization model to deal with the pricing-capacity decisions, using an underlying state-dependent quasi birth–and–death process. The state-dependent feature is essential for modeling the effects of dynamic system congestion and the nonlinearity of waiting. However, it also makes the model quite complicated and unlikely to yield an analytical closed-form solution. Hence, we develop a solution procedure that can efficiently solve the problem numerically using the finiteness property of real scenarios. After obtaining the steady state distribution, we formulate and evaluate the key performance measures of optimal policies under four commonly used competitive strategies, each with a set of operational policies: (1) a marketing strategy that strives to maximize revenue; (2) an operations strategy that tries to minimize the total cost; (3) a joint strategy that maximizes profit by balancing the cost and revenue from both the operations and marketing points of view; and (4) a fixed price strategy that maximizes the profit but has only capacity as a lever.

The contributions of this work are threefold. First, the often-ignored system dynamics are explicitly addressed by an underlying Markov process. By directly taking the dynamic congestion into account, we are able to develop more accurate control policies for demand and capacity. Such realistic treatment provides managers with more powerful tools helping their pricing and staffing decisions.

Second, our paper contributes to the literature on demand and capacity management in the service systems with an integrated marketing–operations perspective. Based on queuing optimization models with a demand rate function that incorporates both price and waiting time sensitivity, we investigate the relative performance of several competitive strategies and their impact on the operating policies. Such an approach is of academic as well as practical interest.

Third, we apply our model to an actual restaurant in assessing alternative competitive strategies for the managers, as well as developing an optimal pricing-capacity policy. The results suggest that a joint marketing–operations perspective can result in a superior operating policy that significantly improves the restaurant's profitability and its return on sales. We also provide the manager with an alternative strategy under pricing restriction, which if properly implemented, can still achieve the chain’s target.

The remainder of this paper is organized as follows. In Section 2, we describe the model and proposes an effective solution procedure. Based on this framework, the optimal pricing-capacity policies under several alternative competitive strategies are formulated in Section 3. In Section 4, we consider a case study applying different strategies to a real local restaurant and provide managerial recommendations. The paper concludes in Section 5.

2. Joint optimization model

In this section, we first formulate the problem of pricing-capacity selection under the joint marketing and operations strategy. We then develop a solution procedure that can yield the optimal policy under the joint strategy, followed by a detailed discussion on each step in the solution procedure. Models under other competitive strategies will be presented in Section 3.

2.1. Formulation and solution procedure

Consider a chain restaurant at certain business hours (e.g., during dinner time) with $s$ tables and the waiting area for at most $r$ parties. Assume the service time for each table follows an exponential distribution with rate $\mu$. Note that the service rate $\mu$ captures the back-house kitchen capacity as well as front-house staff level. Potential parties (customers) form a Poisson stream with exogenous rate $\lambda$. The restaurant serves parties on a first-come-first-served (FCFS) basis. We assume that customers will not renege. Once customers are seated, they will finish their meals. The objective of the manager is to schedule the appropriate staffing levels (and thus service rate $\mu$) and menu price $p$, to optimize system performance under the restaurant chain’s competitive strategy (profit-maximization in this case), with a given service quality guarantee, e.g., 95% of parties wait less than 10 minutes.

The system can be naturally formulated as a $M/M/s/K$ queue, where $K = s + r$. The queueing framework has long been employed to study dynamic service operations (Iittig, 1994, 2002; Khan and Callahan, 1993, and many others). In particular, the $M/M/s/K$ queue properly represents the restaurant system with multiple tables and a waiting area with limited capacity $r$. Restaurants cannot accommodate unlimited demand due to limited kitchen, table, staff and limited waiting room capacity. Customers do not enter when they see a crowded area or a full parking lot. Thus, this framework appropriately models restaurant operations in certain business hours.

Under the joint marketing and operation strategy, the manager's goal is to maximize the average profit. Let $l(x; p, \mu)$ be the arrival rate when there are $x$ parties in the system, and the restaurant is operating at service rate $\mu$ and pricing at $p$. Denote the profit rate function $g(x; p, \mu)$ by

$$g(x; p, \mu) = \lambda(x; p, \mu)p - c_0l(x; p, \mu) + c_1(F(x; s) + R + s),$$

where the first term, $l(x; p, \mu)$, is the revenue rate; the second term, $c_0l(x; p, \mu)p$, is the operating variable cost rate; the third term, $c_1s$, is the labor cost rate, and the last term, $c_1(R + s)$, is the facility cost rate, where $r$ is the waiting space, $s$ is the number of tables, and $\mu$ is a table's service rate. Constants $c_0, c_1$, and $c_2$ represent the unit variable cost, labor cost, and facility cost, respectively.

The optimal pricing-capacity policy $(p^*, \mu^*)$ under a joint strategy can be obtained by solving:

$$\max_{(p, \mu)} G(p, \mu) = \sum_{x=0}^{K} \psi(x; p, \mu)g(x; p, \mu),$$

s.t.

$$F(t) = P(W \leq t) \geq \alpha,$$

$$p \in [\bar{p}, \bar{p}],$$

$$\mu \in [\underline{\mu}, \bar{\mu}],$$

where $\{\psi(x; p, \mu): x = 0, 1, \ldots, K\}$ is a steady state distribution of $X$. Eq. (3) is the waiting time constraint, which requires that the probability of waiting time $W$ less than $t$ should be larger than the pre-specified service level, $\alpha$. $F(\cdot)$ is the probability distribution function of the random variable $W$. The constraint (4) models limited menu prices that the manager can take. This may be determined by the quality of food or required by the restaurant chain headquarters. The constraint (5) represents the possible service rate at which...
the restaurant can operate, mainly determined by the staffing level. If we let \( \Omega \) be the feasible policy space defined by the constraints (3)-(5), it is an uncountable set in theory. However, \( \Omega \) can be well approximated as a finite set in reality.

This formulation is a nonlinear optimization problem with an underlying queuing structure along with a state-dependent demand rate function, \( \lambda(x; p, \mu) \), and a profit rate function, \( g(x; p, \mu) \). Since the feasible policy space \( \Omega \) is finite in reality, we propose the following computation procedure for obtaining the optimal policy by iteratively evaluating and comparing each policy \( (p, \mu) \in \Omega \):

1. **Estimate the demand rate function**, \( \lambda(x; p, \mu) \):
   \[ \lambda(x; p, \mu) = \lambda_0 - \beta p, \]
   \[ \lambda(x; p, \mu) = \lambda_0 e^{-\beta p}, \]
   \[ \lambda(x; p, \mu) = \lambda_0 p^\delta, \]
   where \( \lambda_0 \) is the potential exogenous demand rate. All three functions of \( \lambda(p) \) are decreasing in price \( p \). Such sensitivity also suggests that pricing is an effective level for demand management.

2. **Determine the expected waiting**, \( E[W|X] \), of the current customer, which is determined by the current congestion state, \( x \) (i.e., number of customers in the restaurant). In the \( \text{M}/\text{M}/s/\text{K} \) queue, it is given by:
   \[ E[W|X] = \begin{cases} 0, & \text{if } x < s, \\ \frac{x-1}{\mu}, & \text{if } x \geq s. \end{cases} \]

Compared to the long-term effect of pricing (i.e., \( \lambda(p) \) is independent of the system state \( x \)), service sensitivity is more likely to be an individual experience. To account for this short-term effect, we propose to utilize state-dependent functionals to model the effective demand, \( \lambda(x; p, \mu) \), as

\[ \lambda(x; p, \mu) = \lambda(p) h(E[W|X]), \]

where \( \lambda(p) \) is one of the forms in (6)-(8), and \( h(\cdot) \) is a decreasing function with a value in \([0.1]\) representing the sensitivity of customer waiting. Specifically, \( h(y) \) expresses the percentage of customers who are willing to wait more than \( y \) minutes. This approach takes into the consideration both long-term price sensitivity from the marketing point of view by \( \lambda(p) \) and short-term service (waiting) sensitivity from the operations perspective by \( h(E[W|X]) \).

### 2.3. Steady state distribution

Since demand is influenced by the restaurant performance, the operation of the restaurant under a given pricing-capacity policy is modeled as a state-dependent \( \text{M}/\text{M}/s/\text{K} \) queue. For expositional simplicity, the notation, \( \psi \), is used to denote \( \psi(x; p, \mu) \), the steady state distribution for a given policy \( (p, \mu) \). The distribution \( \psi \), which can be viewed as the fraction of time that the system is in state \( x \), is given as:

\[ \psi_0 = \sum_{k=0}^{s-1} \frac{\lambda_0^{k-1} \mu^k}{k!} \psi_0, \quad 1 \leq x < s, \]

\[ \psi_k = \frac{\lambda_0^{s-1} \mu^s}{s!} \psi_0, \quad s \leq x \leq K, \]

### 2.4. Waiting time distribution

To model the service sensitivity of demand as well as to specify the service quality constraint, we need to have the waiting time distribution, \( F(t) = \text{Pr}[W > t] \). Assuming a FCFS service rule, \( W \) is then a mixed-type random variable, discrete at 0, and continuous at nonzero points. Let \( q_s \) denote the conditional probability that \( x \) customers are in the system given that an arrival is about to occur; this probability is different from \( \psi \), the unconditional probability of \( x \) customers in the system at any arbitrary time point, since the effective arrival process is not Poisson any more. The distribution \( \{q_s\} \) is given by:

\[ q_s = \frac{1}{1 - \lambda_0 x} \psi_x, \quad x \leq K - 1. \]

Consequently, we can obtain \( F(t) \) as follows.

\[ F(0) = \text{Pr}[W \leq 0] = \text{Pr}[\text{system empty at an arrival}] = q_0, \]

\[ F(t) = \{W \leq t\} = 1 - \sum_{k=1}^{s-1} \psi_k \sum_{l=0}^{s-1} \frac{(s)_l}{l!} e^{-\lambda_0 t} \]

### 3. Competitive strategies

In this section, we provide a framework for comparing the performance of four alternative competitive strategies. The first is a cost minimization strategy, called operations strategy (OS) in this paper, in which a manager tries to minimize total cost by adjusting capacity and prices. The second is a revenue maximization strategy, called marketing strategy (MS), in which a manager tries to maximize revenue minus variable cost by increasing capacity to attract customers and manipulate prices. The third is a joint strategy (JS) in which a manager tries to maximize profit by balancing revenue and cost from both the marketing and the operations perspectives. The fourth is a fixed price strategy (FS) in which a manager tries to maximize profit without changing price \( p_0 \).

Mathematically, the strategies under consideration can be modeled through different objective functions. Depending on competitive strategies, different optimal pricing-capacity policies arise. Specifically, the optimal pricing-capacity policy under each strategy is obtained through the following models:
OS: \[ \min_{(p, \mu) \in \Omega} \sum_{x=0}^{K} \psi(x; p, \mu) \{ c_0 \lambda(x; p, \mu)p + c_s s\mu + c_r (r + s) \}, \] (17)

MS: \[ \max_{(p, \mu) \in \Omega} \sum_{x=0}^{K} \psi(x; p, \mu) \{ \lambda(x; p, \mu) \cdot (1 - c_0) \cdot p \}, \] (18)

JS: \[ \max_{(p, \mu) \in \Omega} \sum_{x=0}^{K} \psi(x; p, \mu) \{ \lambda(x; p, \mu) \cdot (1 - c_0) \cdot p - c_s s\mu - c_r (r + s) \}, \] (19)

FS: \[ \max_{\mu \in \Omega} \sum_{x=0}^{K} \psi(x; p_0, \mu) \{ \lambda(x; p_0, \mu) \cdot (1 - c_0) \cdot p_0 - c_s s\mu - c_r (r + s) \}, \] (20)

where set \( \Omega \) is the feasible policy space and \( p_0 \) is a fixed price. Note that the above four models share the same underlying queueing structure, and hence the solution procedure developed in Section 2 applies, with STEP 4 replaced by a computation corresponding the objective function.

4. A casual restaurant case

In this section, we use a real restaurant to show how our models can be applied to derive an optimal operating policy for that restaurant. In particular, we give the detailed procedure of estimating the demand rate function. We then use this framework to compare the effects of alternative competitive strategies on several key system metrics, as well as on their corresponding optimal policies. We also provide recommendations for the manager for staffing and pricing policies to meet the chain’s target.

4.1. Background information

A restaurant in our community was observed between 4:30 pm and 11:00 pm on weekend nights in December, 2004. The data from five Fridays and Saturdays, after excluding special holiday nights, showed very similar patterns in the arrival rate and party size distributions and are used for this research. Furthermore, the restaurant’s operation reports were used and the general manager of the restaurant was consulted frequently to estimate values of parameters used in our model. The restaurant has a waiting room for 11 parties, and its dining room has 54 tables. However, in our analysis, the number of tables is scaled to 40 with a consistent table size (4-top tables), without considering varying sizes and any table combinations. Thus, the assumption is that each party sits at one table without allowing any combination of tables depending on party sizes. During dinner time, the arrival rate is 35 parties per hour, and the average check per party is $41. The service rate per table, \( \mu_0 \), is 0.75 parties per hour. Currently, the restaurant employs 18 servers and 5 cooks. Based on the current operation, we estimate the coefficients of variable cost, labor cost and facility cost are \( c_0 = 0.34, c_s = 7.67, \) and \( c_r = 7.74 \), respectively. The detailed discussions on data collection and cost determination are given in Hwang (2005). The feasible policy space \( \Omega \) is then specified by a price constraint \( p \in [30,100] \); a staffing constraint, which can be transformed into a corresponding service rate constraint, \( \mu \in [0.675,1.425] \); and an overall service quality constraint, i.e., more than 95% of customers wait less than 16.25 minutes. Note that 16.25 minutes is the threshold level when customers begin to be dissatisfied and would balk. See Fig. 1 for the survey data showing how customers wait as their waiting times increase. In order to meet the service standard, the price and service rate should lie within the set \( \Omega \).

4.2. Estimation of the demand process

The exponential form (7) was chosen for the estimation of the demand in this study. The exponential form was also selected over the others by Ittig (1994), who argues that the exponential is realistic at low price and low waiting time. For the determination of the two parameters, \( A_0 \) and \( \beta \), in the exponential model (7), at least two data points are required. One point represents a current price and arrival rate of a restaurant chosen for this study. Each party spends $41 on average, and the average arrival rate during a typical Friday evening is 35 per hour. This point is denoted as \((p_1, A_1)\). The second point is difficult to obtain because managers, in reality, would not be willing to change prices to test a change in demand.

![Fig. 1. Sensitivity of customer waiting.](image-url)
Thus, the second point \((p_2, A_2)\) is approximated from transaction data as \((150, 5)\).

From Eq. (7):

\[
\begin{cases}
A_1 = A_0 \cdot e^{-\beta p_1}, \\
A_2 = A_0 \cdot e^{-\beta p_2},
\end{cases}
\]

\[
\beta = -\left(\frac{\log A_1 - \log A_2}{p_1 - p_2}\right),
\]

\[
A_0 = \exp\left\{\log A_1 - \left(\frac{\log A_1 - \log A_2}{p_1 - p_2}\right) \cdot p_1\right\}.
\]

Using two points \((p_1, A_1) = (41, 35)\) and \((p_2, A_2) = (150, 5)\), the following effective demand with respect to price sensitivity is obtained:

\[
A(p) = A_0 \cdot \exp(-\beta \cdot p) = 72.77 \times \exp(-0.0179 \times p).
\]

Next, to model the impact of the waiting time \(\mathbb{E}[W|x]\) on the demand, the survey data [Hwang, 2005] are used to fit \(h(\cdot)\) in (10). The survey data about customers acceptable waiting times were collected through 810 responses for the question, ‘how long can you wait to be seated before you become dissatisfied and want to leave in a restaurant?’ We assumed that customers balk after the certain amount of time that they were willing to wait. From the logic, the tolerable waiting times constitute the demand curve. From the standard regression (see Fig. 1), we obtain

\[
h(\mathbb{E}[W|x]) = \exp\left(-\frac{(\mathbb{E}[W|x] - 10)^2}{2 \cdot .22^2}\right), \quad \mathbb{E}[W|x] \geq 16.25 \text{ minutes},
\]

where 16.25 minutes is the threshold level when customers begin to balk.

Combining above equations, the demand function from (10) assumes the following form. This is also graphically shown in Fig. 2.

\[
\lambda(x; p, \mu) = \begin{cases} 
72.77 \cdot \exp\left(-0.0179p\right), & \mathbb{E}[W|x] < 16.25 \text{ minutes}, \\
72.77 \cdot \exp\left(-0.0179p - \frac{(\mathbb{E}[W|x] - 10)^2}{2 \cdot .22^2}\right), & \mathbb{E}[W|x] \geq 16.25 \text{ minutes,}
\end{cases}
\]

where \(\mathbb{E}[W|x]\) is given by

\[
\mathbb{E}[W|x] = \begin{cases} 
0, & \text{if } x < s, \\
\frac{1}{(\mu - s)^2}, & \text{if } x \geq s.
\end{cases}
\]

### 4.3. Comparison of competitive strategies

Being in a restaurant chain, the manager needs to meet the chain’s performance target, a return on sales target of 13% with a common service quality level specified by the firm, i.e., more than 95% of customers wait less than 16.25 minutes. Under the current capacity-pricing policy, the restaurant yields a 12.7% return on sales, below the firms required target of 13%. The manager is eager to meet the firm’s target by improving staffing and pricing decisions. The competitive strategies under consideration, as discussed in Section 3, are operations strategy, marking strategy, joint strategy, and fixed price strategy. Under each strategy, the manager wants to obtain the optimal pricing-capacity policy as well as their performance in terms of (1) revenue, (2) total cost, (3) profit, and (4) return on sales.

Using the framework developed in Sections 2 and 3, along with the specification of system parameters in Section 4.1 and the demand rate function in Section 4.2, we numerically solve the problem and summarize the results in Table 1. Fig. 3 shows the contour plots of profits under different strategies and Fig. 4 shows potential profits with respect to different service rate \(\mu\) and price \(p\).

From the computational results, several statements can be made. First, the operations strategy, \((\mu, p) = (0.675, 100)\), minimizes total cost, including variable, labor and facility cost. The results show that the operations strategy tends to have a lower service rate in order to reduce the labor cost, with \(\mu = 0.675\) as the boundary service rate in this case. Furthermore, if the only goal is cost reduction, the manager may well behave in a myopic way and price extremely high to further reduce the demand volume in order to reduce the cost. As a result, the revenue decreases to the lowest
Table 1
Comparison of capacity management strategies.

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Optimal policy</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Service rate</td>
<td>Price ($)</td>
</tr>
<tr>
<td>Current</td>
<td>0.75</td>
<td>41</td>
</tr>
<tr>
<td>Operations</td>
<td>0.675</td>
<td>100</td>
</tr>
<tr>
<td>Marketing</td>
<td>1.425</td>
<td>56</td>
</tr>
<tr>
<td>Joint</td>
<td>0.975</td>
<td>56</td>
</tr>
<tr>
<td>Fixed Price</td>
<td>0.975</td>
<td>41</td>
</tr>
</tbody>
</table>

Fig. 3. Comparison of strategies.

Fig. 4. Profits based on prices and service rates.
one, $916.35. This strategy provides almost no profit and its return on sales, 0.3%, is far below the target. This confirms that, without proper incentive and competitive strategy, restaurant managers may adopt local optimal policies detrimental to the chain.

The marketing strategy \((\mu, p) = (1.425, 56)\) maximized revenue, giving the highest, $1488.90. The marketing strategy tends to have a larger demand. However, it also tends to use over-capacity in order to attract more customers with \(\mu = 1.425\) as the highest service level. Over-capacity increases the total cost to the highest, $1337.90, due to the high service level. As a result, the marketing strategy has almost the same profit as the current strategy ($150.99 \approx 148.86\), whereas the return on sales is less than the current strategy (10.1%/12.7%). Again, with a pure marketing strategy, the manager cannot achieve the target of 13%.

However, if the manager adopts the joint strategy that maximizes profit by balancing cost and revenue from both the operations and the marketing points of view, the joint strategy yields the global optimal, with \((\mu, p) = (0.975, 56)\). Under this strategy, even though the total cost is higher than for the operations strategy ($1167.20 > 913.22\), and the revenue is less than that for the marketing strategy ($1392.70 < 1488.90\), the joint strategy achieves the best trade-off between cost and revenue, with the highest profit, $225.50, and highest return on sales, 16.2%. If the manager's estimation of the demand function is appropriate and if the joint strategy is well implemented, the restaurant can significantly improve profit and return on sales, 16.2%. If the manager's estimation of the demand function is inaccurate or if the joint strategy is not well implemented, the restaurant may experience a situation worse than the current strategy. In fact, neither can improve the return on sales. As for the manager, depending on the restriction of pricing range, he should adopt the optimal policy under either the joint strategy or the fixed price strategy, as both can achieve the chain’s target.

Because various parameters are uncertain, the sensitivity analysis of key parameters is performed. Table 2 shows the application of the originally recommended policy when the values of \(A_0\), \(\beta\), \(c_0\), \(c_s\), and \(c_r\) are over- or under-estimated by 10%. When \(A_0\) is increased or \(\beta\) is decreased by 10%, demand is increased, both the profit and return on sales significantly increase to around $350 per hour and 21%, respectively. On the other hand, the decrease of \(A_0\) or increase of \(\beta\) substantially diminishes the profit and return on sales. That is, the accurate estimation of the customer demand is important to design a good management strategy. The effects of cost parameters such as \(c_0\), \(c_s\), and \(c_r\) are much smaller. Even if the estimation is somewhat inaccurate the optimal control policy still generates reasonably good performance.

5. Conclusions and future study

In this paper, a queueing-based optimization model is developed to address the problem of demand and capacity management in the restaurant industry. The model explicitly captures the long-ignored system dynamics (congestion) and the delicate interaction

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Change (%)</th>
<th>Hourly profit ($)</th>
<th>Return on sales (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No change</td>
<td></td>
<td>225.5</td>
<td>16.2</td>
</tr>
<tr>
<td>(A_0) +10%</td>
<td></td>
<td>348.8</td>
<td>21.1</td>
</tr>
<tr>
<td>(A_0) –10%</td>
<td></td>
<td>150.8</td>
<td>11.2</td>
</tr>
<tr>
<td>(\beta) +10</td>
<td></td>
<td>159.8</td>
<td>11.7</td>
</tr>
<tr>
<td>(\beta) –10</td>
<td></td>
<td>359.7</td>
<td>21.6</td>
</tr>
<tr>
<td>(c_0) +10</td>
<td></td>
<td>174.5</td>
<td>12.5</td>
</tr>
<tr>
<td>(c_0) –10</td>
<td></td>
<td>276.8</td>
<td>19.9</td>
</tr>
<tr>
<td>(c_s) +10</td>
<td></td>
<td>191.3</td>
<td>13.7</td>
</tr>
<tr>
<td>(c_s) –10</td>
<td></td>
<td>260.3</td>
<td>18.7</td>
</tr>
<tr>
<td>(c_r) +10</td>
<td></td>
<td>186.0</td>
<td>13.4</td>
</tr>
<tr>
<td>(c_r) –10</td>
<td></td>
<td>264.9</td>
<td>19.0</td>
</tr>
</tbody>
</table>

*Fig. 5. Suboptimal strategy under fixed price.*
between system performance and demand on a real-time basis by an underlying quasi birth-and-death process along with state-dependent demand and objective functions. With this framework, we then examine the relative performance of the optimal pricing-capacity policies under four commonly used competitive strategies. Finally, the case of a local casual restaurant is studied and the results illustrate the immediate applicability of our model.

In practice, managers often focus on only one aspect of the demand and capacity problem. They tend to either adopt the marketing point of view and provide unnecessarily superior service quality that customers may not recognize; or from an operations point of view, strive to cut the cost by lowering service level to the bare minimum with the least possible capacity. However, our study shows that neither strategy is ideal for most casual restaurants with the goal of profit-maximization. Instead, a joint strategy that balances both marketing and operations perspectives should be embraced. The result indicates that both efforts to increase service quality and reduce cost can be balanced; resulting in maximum profit in services. This study confirms that improving service quality is not always contradictory to reducing costs as found in other studies (Anderson et al., 1997; for example). Thus, our work suggests that interdisciplinary perspectives should be applied in research on capacity planning decisions as well as in the practice of the restaurant industry. Our work also indicates that companies need to provide proper incentives for the restaurant managers to choose appropriate capacity and pricing strategies that are consistent with company's strategic goal. Without appropriate incentives in place, the managers may choose capacity and pricing policies that can be detrimental to the company.

For the future research, more avenues lie ahead. From a modeling perspective, one avenue would be to capture the complexity of restaurant operations, especially the multi-stage feature involved in customer dining experience. Thus multi-stage tandem queues or queueing networks should be utilized to reflect more relevant details of the restaurant system. Such complicated framework may be hard in yielding simple form solutions. Nonetheless, more accurate staffing and pricing control policies can be examined under this more precise framework. Also, from a practical perspective, empirical investigations of the effects of the price sensitivity, as well of the waiting time sensitivity on demand should be carried out in an experimental environment. Such studies will be of great value to restaurant managers, as well as to the applicability of our model.

References