Article

Cornell Hospitality Quarterly
2019, Vol. 60(4) 378–388

DOI: 10.1177/1938965518805685
journals.sagepub.com/home/cqx
SAGE

The Value of Timing Flexibility in Restaurant Reservations

Gary M. Thompson[1]

Abstract

A recent paper presented and evaluated 10-integer programming models for restaurant reservations, finding that the models that pooled reservations by same-size tables were superior to models that matched reservations to specific tables. An assumption in all the models was that demand timing was inflexible and the evaluation of the models assumed that all customers arrived exactly at the designated reservation time. Although restaurant customers may have an ideal dining time, many customers have some flexibility in when they would accept a reservation. To address this, we extend the pooled-table models to allow for demand timing flexibility and evaluate the models with a range of differences between customer arrival times and their designated reservation time. With the highest flexibility in demand timing, a top-performing model increased revenue by over 21% compared with rigid demand timing. Fortunately for restaurateurs, the increased revenue came at the expense of only a small deterioration in customer service, as measured by the percentage of parties that need to wait for a table upon arrival.

Keywords

revenue management, services, restaurants, integer programming applications

Introduction

Despite there being a large body of work on restaurant revenue management, restaurant reservations have seen little attention. Given the size of the restaurant industry, for which sales were forecast to exceed US$552 billion in 2017 (Anonymous, 2017), the lack of attention is surprising. Compared to using walk-in customers, reservations can offer restaurants the opportunity to plan for the number and timing of parties. A big driver for filling a reservation book is maximizing revenue. Revenue can be increased by accepting more parties through careful scheduling of reservations, taking higher value customers, or both. A challenge arises, however, in dealing with the variation in dining duration, an issue first identified by Kimes, Chase, Choi, Lee, and Ngonzi (1998). If one uses only the mean dining time, for example, there is a notable risk that tables will not be ready when customers arrive for their reservations (Thompson, 2015b). The issue becomes, then, one of a trade-off between revenue and customer service. Customers expect their tables to be ready (Kimes, 2008) and are not easily pacified when they are not (McDougall & Levesque, 1999). So, while lower levels of customer service (i.e., more parties needing to wait for a table) can appear to yield higher revenue, that really is a false revenue, because it is unlikely to be sustainable.

In a recent study, Thompson (2015b) presented and evaluated 10 optimization models for restaurant reservations. Decisions in the models were inward oriented—the mix of tables—and outward oriented—the reservation slots that would be made available to customers. A limitation of all the models is that demand was treated as static and the models were evaluated assuming all customers arrived at their designated reservation time. Although restaurant customers may have an ideal dining time, many customers have some flexibility in when they would accept a reservation. For example, a search for restaurant reservations on the popular restaurant reservations portal OpenTable (2018) shows that available dining times are listed within two or more hours of the requested time. The restaurant reservations portal Resy (2018) does not request a desired dining time, but simply lists all the available reservation slots for the day. This listing of all times suggests that guests will pick the option that best meets their needs from the available reservation slots.

To address the issue of demand timing flexibility, we extend some of the models presented by Thompson (2015b) and use these models to evaluate the benefits of considering demand timing flexibility. We perform the evaluation using

[1]Cornell University, Ithaca, NY, USA

Corresponding Author:
Gary M. Thompson, Professor of Service Operations Management, School of Hotel Administration, Cornell University, 352 Statler Hall, Ithaca, NY 14850, USA.
Email: gmt1@cornell.edu

variation between customers' actual arrival times and their designated reservation times. Compared to rigid demand timing, we find that a top-performing model increased revenue by over 21% when demand was completely flexible. Fortunately for restaurateurs, the increased revenue comes at the expense of only a small deterioration in customer service, as measured by the percentage of parties who need to wait for a table upon arrival.

The structure of the remainder of the article is as follows. We review relevant literature; present our models of restaurant reservations with flexible demand timing, describe the variants of the models that we evaluate, describe the test scenarios we created to evaluate the models, and present results. We close with a discussion. An Online Appendix shows three models for a simple restaurant reservations context.

Literature Review

Kimes et al. (1998) were the first to coin the term "restaurant revenue management." Most of the significant body of work since then has been related to restaurants taking walk-in customers. Thompson (2010) provides a summary, and identified reservations as one of the little-addressed topics. Seven studies are of notable relevance to our current work: one experiment-based study, one survey-oriented study, and five modeling studies.

McDougall and Levesque (1999), examined, via an experiment, attempts to diminish the effects of a table that was not ready at the time of a reservation. The mitigation efforts they examined were apologies, assistance, compensation, and assistance plus compensation. None of these efforts led to positive future intentions toward the restaurant. In a convenience survey of restaurant customers, Kimes (2008) found that customers had high expectations that tables would be ready at the time of the reservation. The consequence of these two studies is that a restaurant must honor reservations, at their designated times. This yields the key customer service metric in our study: the percentage of parties that need to wait for a table upon arrival.

The first modeling investigation into restaurant reservations was conducted by Bertsimas and Shioda (2003). They considered rolling decisions, with the choice to accept or decline customers with reservations and walk-in customers. Given the negative implications of customers having to wait for a promised reservation, from the studies noted earlier, turning away customers with reservations would create serious problems for restaurateurs.

Thompson and Kwortnik (2008) examined reservations in a dining context like a cruise ship, where customers were seated in large batches to a predetermined mix of tables. They focused on whether reservations should be linked to specific tables or pooled by table size. In a survey they conducted, four fifths of restaurants reported matching specific tables to reservations. By contrast, they observed that pooling reservations by table size yielded fewer parties waiting for a table at the reserved time.

The decision of whether or not to take reservations was examined by Alexandrov and Lariviere (2012). They evaluated strategies to alleviate the effect of no-show parties, which was a key factor in their models. Their models did not distinguish between party sizes and considered capacity in aggregate, instead of being time dependent. Our investigation assumes that reservations will be taken. Time-dependent capacity usage is important for our investigation, as are differences in party sizes, both in the amount of capacity (i.e., the size of tables) and the duration the capacity is in use.

Guerriero, Miglioni, and Olivito (2014) presented several models for strategic and operational revenue management decisions in restaurants. Their TMP_3 and TMP_4 models dealt with a problem somewhat like what we examine here, in that their models determined an ideal mix of tables and determined which reservation slots would be open to customers. Both models pooled tables of given sizes. TMP_3 required that tables not be combined, as we assume here, whereas TMP_4 allowed tables to be combined to seat larger parties. Both models allowed the number of different size tables to change during the meal period, which is impractical. TMP_4 also failed to consider the time required to combine tables and whether the tables to be combined were proximal to each other. Finally, they made no attempt to evaluate how well a recommended solution might play out, given variation in dining durations. We note, however, that they did consider offering parties the option of dining later, if their desired dining time was not available, with no limits on the delay time.

Notably relevant to our work is Thompson (2015b), who presented and evaluated 10 different integer programming models for restaurant reservations. The models optimized the mix of capacity and the reservations to make available to customers, and distinguished between parties based on their size. Reservations were staggered and overlapped. There were five models each of two types, one where tables were pooled by size and the other where reservations were matched to specific tables. Both types of models were implemented with fixed dining durations and with flexible dining durations. All five of the pooled-table models fell on the service-level-revenue pareto frontier, as did two of the matched-table models. The matched-table models proved much more difficult to solve. The best performing model in the study was a fixed-dining-duration, pooled-table model, where the dining duration for each size party was rounded-up to the next higher number of 15-minute periods, plus one period. The results of Thompson's (2015b) study supported Kimes et al.'s (1998) contention that reducing variation in dining duration would better enable restaurateurs to determine the timing of reservations.

A limitation of Thompson's (2015b) models is the rigid timing of customer demand. Relaxing that assumption is the basis for our study, which allows us to examine the effects of demand timing flexibility. Another assumption of Thompson's (2015b) study was that all customers arrived at their designated reservation time. We relax that assumption in the current study.

Integer Programming Models for Restaurant Reservations With Flexible Demand

We use five of the pooled-table, pareto-optimal models presented by Thompson (2015b), extended to handle flexible demand, supplemented with a model for handling fully flexible demand. Below, we define sets and indexes, parameters, and decision variables. Most of these were defined in Thompson (2015b); we note the components new to our models.

Indices and sets:

c = index for party sizes;
i = index for table seats;
j = index for specific tables;
l, l' = index for dining length, in periods;
p, p' = index for periods;
m = index for moved periods for demand (*new*);
C = set of customer (party) sizes;
P = set of dining periods in which reservations are accepted;
S = set of table sizes; and
L_c = set of possible meal durations (in periods) for a party of c people.

Parameters:

dmd_{cp} = demand for reservations (in number of parties) of size c at period p;
$lenmin_c$ = minimum meal duration (in periods) for a party of c people;
$lenmax_c$ = maximum meal duration (in periods) for a party of c people;
$lenprop_{cl}$ = probability that a party of c people will take longer than l periods to dine;
$mxtbls_i$ = maximum number of individual tables of i seats = $\left\lfloor spaceavl / space_i \right\rfloor$;
$mxsftpr$ = maximum number of periods that demand can be shifted (*new*);
$mxsft_m$ = maximum proportion of demand in a period that can be shifted $\pm$ m periods, where $m \leqslant mxsfrtr$ (*new*);
$mxsftres_c$ = maximum number of reservations that can be shifted for a party of size c (*new*);
$spaceavl$ = space available for seating in the restaurant;
$space_i$ = space required for a table of i seats; and
$value_c$ = value of a party of c people.

Decision variables:

$a_{cpp'}$ = number of reservations for a party of size c, preferred in period p, but accepted in period p' (*new*);
n_i = number of tables with i seats; and
g_{clpi} = number of reservations accepted for a party of size c, length (duration) l, at period p, placed in a generic (pooled) table with i seats.

The basic pooled-table model presented by Thompson (2015b), which he called TP1, is

$$MAX \sum_{c\in C} value_c \left(\sum_{l\in L_c} \sum_{p\in P} \sum_{i\in S} g_{clpi} \right). \tag{1}$$

$$\sum_{i\in S} space_i * n_i \le spaceavl \tag{2}$$

$$\sum_{\{c\in C|c\le i\}} \sum_{l\in L_c} \sum_{\{p'\in P|p-l+1\le p'\le p\}} g_{clp'i} \le n_i \; \forall \; p\in P, \; i\in S \tag{3}$$

$$\sum_{l\in L_c} \sum_{\{i\in S|i\ge c\}} g_{clpi} \le dmd_{cp} \; \forall \; c\in C, \; p\in P, \tag{4}$$

$$n_i = \{0,1,2,\ldots,mxtbls_i\} \; \forall \; i\in S. \tag{5}$$

$$\begin{aligned} g_{clpi} = \{0,1,2,\ldots,dmd_{cp}\} \; \forall \; c\in C, \; l\in L_c, \\ p\in P, \{i\in S | i\ge c\} \end{aligned} \tag{6}$$

The objective expressed in Equation 1 maximizes the total value of the served parties. Equation 2 ensures that the space used for all tables does not exceed the available space. Equation 3 requires that the number of parties served, by period and table size, does not exceed the number of appropriate-sized tables. Equation 4 limits the number of parties served to the available demand. Equations 5 and 6 impose the integrality restrictions for the number of tables and parties accepted. We have modified Equations 3, 4, and 6 from those presented by Thompson (2015b), to ensure parties can be seated only in tables with sufficient seats, an apparent oversight in the presentation of his models.

Thompson (2015b) implemented TP1 using a fixed dining duration for each party size, which resulted in the sets L_c each having only a single member. He examined TP1(0), TP1(1) and TP1(2), which set the dining duration to the rounded-up number of periods plus zero, one, and two periods, respectively. These three variants, which we also consider, all fell on the revenue-service-level pareto frontier.

The second pooled-table model presented by Thompson (2015b), which he called TP2, is the combination of Equations 1, 2, 3, 4, 5, 6, and

$$\sum_{\{l' \in L_c | l' \geq l\}} g_{cl'pi} \geq lenprop_{cl} \sum_{l' \in L_c} g_{cl'pi} \; \forall c \in C, \quad \{l \in L_c \mid l > lenmin_c\}, p \in P, \{i \in S | i \geq c\}. \tag{7}$$

The additional constraint set expressed in Equation 7 in TP2 ensures that a sufficient proportion of longer duration parties are included in the accepted reservations. We modified this constraint as well, for the reason noted earlier. A model parameter under user control is the size of the Sets L_c, for each party size. Thompson (2015b) limited the size of the sets to one and five periods more than the rounded-up mean dining duration.

TP1Dflex, our first new model, is an extension of TP1. TP1Dflex, which represents demand timing flexibility with new variables, is the combination of Equations 2, 3, 5, and

$$MAX \sum_{c \in C} value_c \left(\sum_{l \in L_c} \sum_{p \in P} \sum_{i \in S} g_{clpi} \right) - 0.01 * \sum_{c \in C} \sum_{p \in P} \sum_{\{p' \in P | p - mxsftpr \leq p' \leq p + mxsftpr\}} |p' - p|. \tag{8}$$

$$\sum_{\{p' \in P | p - mxsftpr \leq p' \leq p + mxsftpr\}} a_{cpp'} \leq dmd_{cp} \; \forall c \in C, \; p \in P \tag{9}$$

$$\sum_{\{p' \in P | p - mxsftpr \leq p' \leq p - m\}} a_{cpp'} + \sum_{\{p' \in P | p + m \leq p' \leq p + mxsftpr\}} a_{cpp'} \leq \lfloor mxshft_m * dmd_{cp} \rfloor \forall c \in C, \; p \in P, \; 1 \leq m \leq mxsftpr \tag{10}$$

$$\sum_{l \in L_c} \sum_{\{i \in S | i \geq C\}} g_{clpi} = \sum_{\{p' \in P | p - mxsftpr \leq p' \leq p + mxsftpr\}} a_{cp'p} \quad \forall c \in C, \; p \in P \tag{11}$$

$$\sum_{p \in P} \sum_{\{p' \in P | p - mxsftpr \leq p' \leq p + mxsftpr\}} a_{cpp'} \leq mxsftres_c \quad \forall c \in C \tag{12}$$

$$g_{clpi} = \left\{ \begin{array}{l} 0, 1, 2, \ldots, dmd_{cp} \\ + \sum_{m=1}^{mxsftpr} \left(\begin{array}{l} mxshft_m * dmd_{c(p-m)} \\ + mxshft_m * dmd_{c(p+m)} \end{array} \right) \end{array} \right\} \quad \forall c \in C, \; l \in L_c, \; p \in P, \{i \in S | i \geq c\} \tag{13}$$

In addition to the revenue originating from the reservations, the objective of TP1Dflex expressed in Equation 8 includes a small penalty for demand being moved, with larger movements having larger penalties. This ensures the solution is parsimonious in its use of demand movements. Equation 9 ensures that the total accepted reservations for each party size, originating from a period, does not exceed the demand for the period and party size. Equation 10 limits the amount that demand can be shifted, by party size, period, and shift amount. The limit is the rounded-down amount of demand that can be moved, shown in the right-hand side of Equation 10. Equation 11 ensures, for each party size and period, that the number of parties served equals the amount of demand across the range of periods from which the demand can be shifted. Equation 12 allows one to limit the total amount of demand shifted to other periods for each party size. This would allow, for example, one to use the model to explore the response curve between revenue and the amount of demand being shifted. Finally, Equation 13 imposes the appropriate integrality restrictions for the reservation variables in this, more flexible model.

Note that when $mxsftpr = 0$, TP1DFlex is identical to TP1. When implementing TP1Dflex, there are the issues of the dining duration inflation and the value of *mxsftpr* used. As such, we use TP1Dflex (DurIncInPrds, *mxsftpr*) to fully describe the model.

The second new model that incorporates demand timing flexibility, TP2Dflex, is based on Thompson's (2015b) TP2 model. It is the combination of Equations 2, 3, 5, 7, 8, 9, 10, 11, 12, and 13.

As with TP1DFlex, TP2Dflex has the issues of the size of the Sets L_c for each party size, so we use TP2Dflex(MaxL$_c$, *mxsftpr*) to fully denote the model.

Our third new model, TP1DFullFlex is the combination of Equations 1, 2, 3, 5, and

$$\sum_{l \in L_c} \sum_{\{i \in S | i \geq c\}} g_{clpi} \leq \sum_{p \in P} dmd_{cp} \quad \forall c \in C. \tag{14}$$

$$g_{clpi} = \left\{ 0, 1, 2, \ldots, \sum_{p \in P} dmd_{cp} \right\} \quad \forall c \in C, \; l \in L_c, \; p \in P, \{i \in S | i \geq c\} \tag{15}$$

Equation 14 requires, for each party size, that the total number of parties served cannot exceed the total demand. Equation 15 imposes the appropriate integrality restrictions for the fully flexible demand. When implementing TP1DFullFlex, there is the issue of the dining duration inflation and so we use TP1DFullFlex(DurIncInPrds) to fully describe the model. The Online Appendix shows TP1(1), TP1Dflex (DurIncInPrds = 1, $mxsftpr$ = 2), and TP1DFullFlex(DurIncInPrds = 1) for a sample problem.

Table 1.
Demand Timing Flexibility.

	Extent of Demand Movement		
Demand Timing Flexibility Level	± 15 Min (1 Period)	± 30 Min (2 Periods)	± 45 Min (3 Periods)
0	0	0	0
1	0.333	0	0
2	0.667	0.333	0
3	1	0.667	0.333

Model Variants Examined

We evaluated three variants of TP1Flex, based on different round-up values, and three variants of TP2Flex, based on different sizes of the L_c sets. The three TP1Flex round-up values were zero, one, and two periods, matching the values used by Thompson (2015b). We used limits of the rounded-up mean dining duration plus one, two, and five periods for the three variants of TP2Flex. Thompson (2015b) used one and five periods only. For TP1DFullFlex we used limits of the rounded-up mean dining duration plus two periods.

For each of our TP1Flex and TP2Flex models, we examined four levels of demand timing flexibility, as summarized in Table 1. This table reports the proportion of the demand that can be shifted. For example, at the highest level of flexibility, one third of demand can be shifted by ± 45 minutes (± three 15-minute periods). At the lowest flexibility, no demand can be shifted.

Test Scenarios

To test the benefits of demand flexibility across the models, we created a set of 3,840 test scenarios based on those of Thompson (2015b), as summarized in Table 2. Below, we give a brief description of the factors and factor levels. Thompson (2015b) provides a rationale for the inclusion of these factors, which we briefly note.

Our three restaurant sizes were 40, 80 and 160 seats. Although the largest size falls between the median and mean number of seats in a survey reported by Thompson (2011a), he included the smaller restaurant sizes because "the smaller pool of available tables would reduce flexibility" (Thompson, 2015b, p. 312). The reduced flexibility was manifested in the models, typically showing the poorest service levels in the 40-seat restaurant, where service was measured as the proportion of customers waiting for a table upon their arrival.

The demand load factor had four levels, from a high of 120% of full capacity to a low of 90% of full capacity. Earlier studies, all for restaurants taking walk-in parties only, used load factors between 90% of full capacity (Thompson, 2015a) and 300% of full capacity (Thompson, 2011b), with the majority around or slightly over full capacity. Thompson (2015b) hypothesized that higher demand loads could lead to fuller restaurants and reduced service levels, which proved to be the case. Demand load comes into play as follows. The mean number of reservations per 15-minute period is equal to the number of seats in the restaurant, times the demand load factor divided by 100, divided by four (to convert to 15-minute periods), divided by the mean party size (described below).

Table 2.
Factors and Factor Levels for the Test Environments.

Factor	Number of Levels: Levels
Restaurant size (seats)	3: 40, 80, 160
Demand load factor (% of full capacity)	4: 90, 100, 110, 120
Day length (hr)	2: 2, 4
Mean party size (people)	2: 2.5, 3.0
Across-party duration variation (ratio of mean duration of a party of 10 people to mean duration for a party of one person)	2: 1.5, 2.0
Within-party duration variation (coefficient of variation)	2: 0.15, 0.30
Average check variation (ratio of spend per person for parties of 10 to the spend per person of parties of one)	2: 0.9, 0.8
Mean arrival time discrepancy (mean of the actual arrival time minus the designated reservation time)	5: –10, –5, 0, 5, 10

Two day lengths were used: 2 and 4 hours, each broken into 15-minute periods. Thompson (2015b) expected that service levels would decline under the longer day length because the longer day would give more timing flexibility and so results in fuller restaurants. In his study, the worst service levels occurred under the longer day length with four of the seven pareto-optimal models; while day length was not a distinguishing factor for the other three.

Mean party size had two levels: 2.5 and 3.0 people per party. Earlier studies reported mean party sizes of 2.5 people (Kimes & Robson, 2004), 2.55 people (Thompson, 2011b), and from 2.5 to 2.9 people across the days of the week (Kimes & Thompson, 2005). Although Thompson (2015a) explicitly stated no expectation about the effect of party size on customer service, the worst service level results for five of the seven pareto-optimal models occurred with the larger mean party size.

It has been observed previously that larger parties take longer to dine (Bell & Pliner, 2003; Kimes & Robson, 2004, and Kimes & Thompson, 2004). There were two levels of the factor that addressed this phenomenon. Across-party duration variation factor had ratios of the dining duration of a party of 10 to the dining duration of a party of one, being

1.5 and 2.0, consistent with Thompson (2015a). Thompson (2015b) did not have any a priori expectations of the effect of party size on service. However, four (one) of the seven pareto-optimal models had the worst-case service levels under the lower (higher) ratio, whereas for two models this was not a distinguishing factor in worst-case performance.

The factor within-party duration variation had two levels, representing coefficients of variation in same-size party dining durations of 0.15 and 0.30, which are within the ranges reported in the literature of 0.16 to 0.50 (Bell & Pliner, 2003; Kimes & Robson, 2004). Thompson's (2015b) expectation, based on the statements of Kimes et al. (1998), was that the higher level of this factor would result in lowered service levels. This was indeed the case in his study for worst-case performance of all seven pareto-optimal models.

Average check variation represented differences in the spend per person across party sizes, as it has been observed that larger parties spend less per person than smaller parties (Kimes & Robson, 2004; Kimes & Thompson, 2004). The two levels had a ratio of the per-person-spend for a party of 10 compared with that of a party of one of 0.8 and 0.9. Thompson (2015b) explicitly stated no expectation about this factor's effect on service levels and it was not a distinguishing factor in the worst-case performance of any of the seven pareto-optimal models in his study.

Despite not being able to find any information on the discrepancy between customers' actual arrival time compared with their designated reservation time, we wished to explore the effect of such differences. We used five levels for the mean arrival time discrepancy: means of −10, –5, 0, 5, and 10. Negative (positive) values indicate arrivals before (after) the assigned reservation time, on average. In all cases, we assumed that the standard deviation would be 3.67 minutes, resulting in 99.7% of parties arriving within plus or minus 11 minutes of the mean discrepancy.

The full-factorial design, with the levels of the eight factors reported in Table 2 yielded 1,920 scenarios. For each of these, as did Thompson (2015b), we created two sets of reservation demand patterns, giving a total of 3,840 test scenarios. TP1DFullFlex and the four demand-shifting versions of the six total variants of TP1Dflex and TP2Dflex were evaluated across all scenarios.

We followed the same procedure for solving the models as described by Thompson (2015b). We developed a model for a particular context, attempted to solve it optimally, and simulated how well the solution would perform. Details on each step follow. To create the model, we used a mathematical programming system file generator we developed in Excel® using a Visual Basic for Applications macro. We then used the Gurobi solver (Gurobi, 2015), run from the command line, with a time limit of 10 minutes per problem. After solving a problem, we loaded the solution and simulated 100 days of that reservation mix using a Visual Basic for Applications macro in Excel®. All of our investigations were performed on an Intel i5-based personal computer, with four cores, which has a Linpack (Dongarra & Luszczek, 2011) benchmark of 1,737.7 MFLOPS. Our procedure for solving TP1DFullFlex was different: We created a model in Excel® and solved it using OpenSolver (OpenSolver, 2018). Like the other models, however, 100 days of the reservation mix was simulated.

The simulation incorporated randomness in the dining durations. Following Thompson (2015b), we used a lognormal distribution for dining times, because the longer right tail matches the extended durations commonly seen in restaurants. Like Thompson (2015b), we assumed that parties would be assigned to the first available table of the appropriate size (from the model solution), and that parties would remain in the same table for the duration of service. In contrast to Thompson (2015b), we did not assume that parties arrived on time. Rather a party would arrive at the designated reservation time, plus the mean arrival time discrepancy, plus a normal random variate with a mean of 0.00 and a standard deviation of 3.67 minutes.

Like Thompson (2015b), from the simulation we collected performance metrics on the average revenue, the average percentage of parties that waited upon arrival for a table, and the average wait time of those parties who waited for a table. Of these metrics, we consider the average percentage of parties who waited for a table on arrival to be the key customer service metric. The rationale for the metric was described earlier: Customers expect their tables to be ready (Kimes, 2008) and are not easily pacified when they are not (McDougall & Levesque, 1999).

Results

The results of our study are presented in Figures 1 and 2 and Tables 3 to 6. Table 3 reports the sizes and solution times of TP1Dflex and TP2Dflex. Across all the models, model size and solution times grows across the levels of demand timing flexibility. In most instances, other than for TP2Dflex using one extra period, average solution times were under 0.5 seconds. As observed by Thompson (2015b) for TP2, solution times for TP2Dflex were faster when the model used more extra periods, despite being larger in size.

Table 4 compares our daily revenue results with no demand timing flexibility to those of Thompson (2015b). As we based our study on the same experimental factors, we expected that the no-flexibility results would closely match. The results match closely for all the models, offering a validation of our results.

Table 5 gives the average daily revenue and percentage of parties waiting for a table, for the variants of TP1Dflex

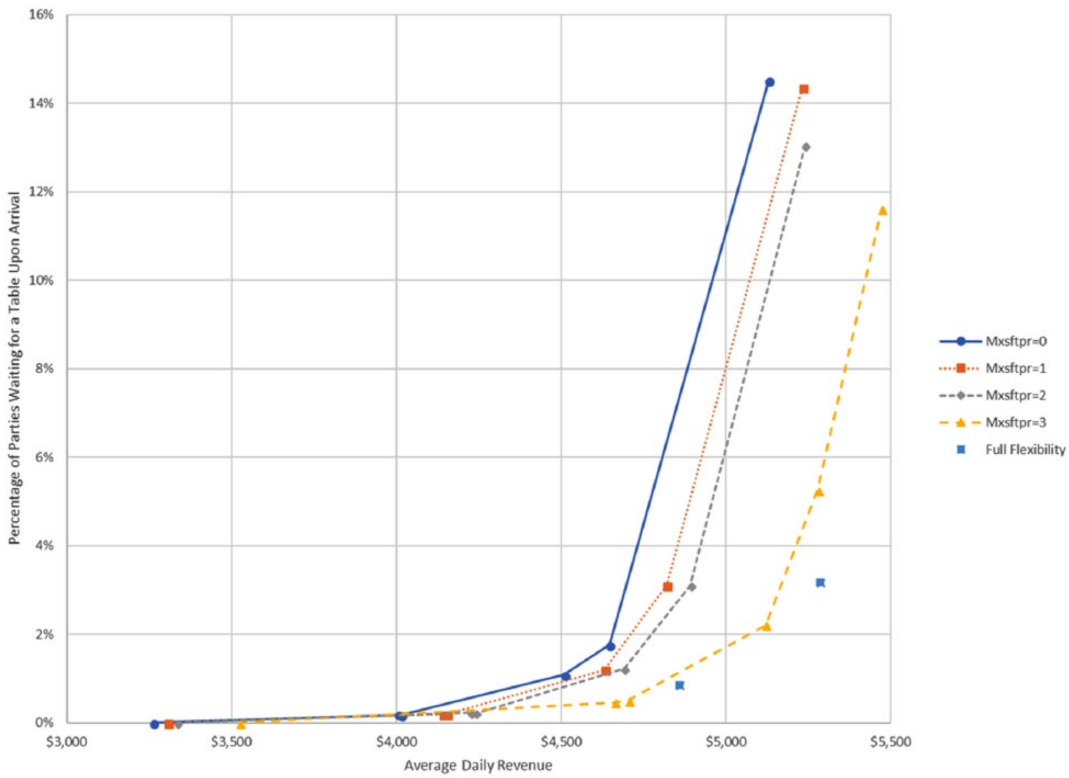

Figure 1.
Relationship between revenue and the percentage of parties that wait for a table upon arrival by model and dining flexibility level for a mean arrival time discrepancy of zero.
Note. From left to right, for each level of demand timing flexibility, the models are TP2Dflex (extra periods = 5), TP1Dflex (DurIncInPrds = 2), TP2Dflex (extra periods = 2), TP1Dflex (DurIncInPrds = 1), TP2Dflex (extra periods = 1), and TP1Dflex (DurIncInPrds = 0).

and TP2Dflex, ordered by increasing revenue, under the case of no demand timing flexibility. The revenue and service differences were large. Comparing the models at the extremes, TP1Dflex (round-up = 0) had 1.58 times the revenue of TP2Dflex (extra periods = 5), but between 4,544 and 5,747 times as many parties waiting for a table across the levels of mean arrival time discrepancy.

Figure 1 shows the relationship between the percentage of parties who wait for a table and revenue, by level of demand timing flexibility, for the models. As expected, increasing the level of demand timing flexibility increases revenue; that is, the curves shift to the right. Also as expected, increasing the level of demand timing flexibility increases the percentage of customers who wait for a table; that is, the same-model comparisons shift upward. Noteworthy are the two models at the pronounced inflection point on the curves—TP1Dflex (round-up = 2) and TP2Dflex (extra periods = 2). These models provide an excellent combination of revenue and customer service.

Table 6 presents the improvement in revenue across the levels of demand flexibility. For all models, the improvement in revenue from no demand timing flexibility to a low level of demand timing flexibility was between 1.4% and 3.7%. For the models that were closest to the pronounced inflection point in the service-level-revenue relationship—TP1Dflex (round-up = 2) and TP2Dflex (extra periods = 2)—the revenue bump of moving from no demand timing flexibility to the highest demand timing flexibility was the greatest, in both cases in excess of a 16% revenue bump. The two models at the extremes of the service-level-revenue pareto frontier—TP2Dflex (extra periods = 5) and TP1Dflex (round-up = 0)—had the lowest revenue bumps from moving from no demand timing flexibility to high demand timing flexibility, under 9% for both models.

Figure 2 shows the relationship between the percentage of parties who wait for a table upon their arrival, across the levels of mean arrival time discrepancy, for the four variants of TP1Dflex (round-up = 2) and TP1DFullFlex (round-up = 2). With demand shifts of two periods or fewer, there was little effect on customer service. Customer service deteriorated somewhat with a

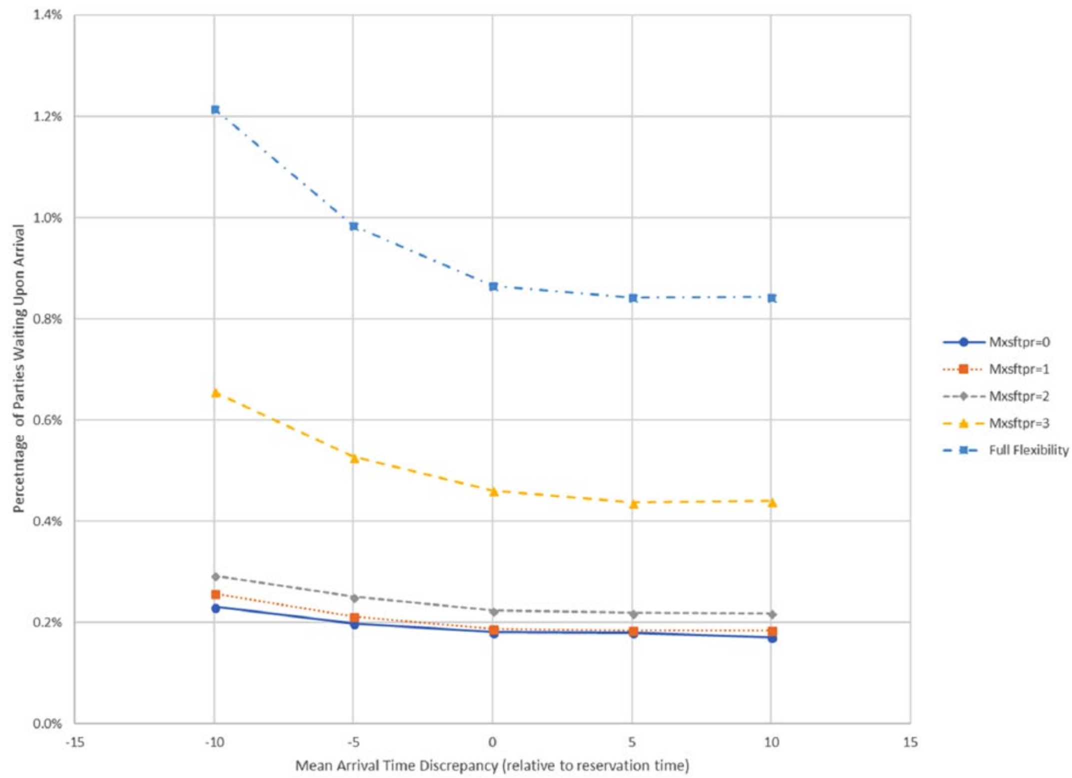

Figure 2.
Relationship between the percentage of parties that wait for a table upon arrival and the mean arrival time discrepancy for TP1Dflex (DurIncInPrds = 2) and TP1DFullFlex(DurIncInPrds = 2).

Table 3.
Model Size and Solution Times by Model and Level of Demand Timing Flexibility.

Model	*Mxsftpr*	Number of Variables	Number of Constraints	Number of Nonzeros	Solution Time (s)
TP1Dflex	0	227.1	181.0	1,348.4	0.09
	1	326.2	308.5	1,437.3	0.19
	2	335.8	333.3	1,519.9	0.21
	3	579.4	333.3	2,443.3	0.31
TP2Dflex (extra periods = 1)	0	671.3	901.0	5,229.5	0.19
	1	770.3	1,028.5	4,874.2	1.93
	2	780.0	1,053.3	4,956.9	3.00
	3	1,169.4	1,053.3	7,089.9	29.28
TP2Dflex (extra periods = 2)	0	893.4	1,261.0	8,106.2	0.11
	1	992.4	1,388.5	7,528.8	0.18
	2	1,002.1	1,413.3	7,611.5	0.24
	3	1,464.5	1,413.3	10,647.7	0.45
TP2Dflex (extra periods = 5)	0	1,351.8	2,057.5	16,027.6	0.06
	1	1,450.9	2,185.0	14,991.8	0.08
	2	1,460.5	2,209.8	15,074.5	0.08
	3	2,092.8	2,209.8	20,944.7	0.13

Note. Mxsftpr = maximum number of 15-minute periods that demand can be shifted.

Table 4.
Average Daily Revenue by Model With No Demand Timing Flexibility From the Current Study and From Thompson (2015b).

Our Results With No Timing Flexibility		Thompson's (2015b) Results		Daily Revenue Discrepancy
Model	ADR	Model	ADR	
TP2Dflex (extra periods = 5)	US$3,261.72	TP2 (extra periods = 5)	US$3,254.72	US$7.00
TP1Dflex (DurIncInPrds = 2)	US$4,005.98	TP1 (DurIncInPrds = 2)	US$4,005.44	US$0.54
TP2Dflex (extra periods = 2)	US$4,013.97	NA	NA	NA
TP1Dflex (DurIncInPrds = 1)	US$4,508.49	TP1 (DurIncInPrds = 1)	US$4,531.54	(US$23.05)
TP2Dflex (extra periods = 1)	US$4,645.86	TP2 (extra periods = 1)	US$4,651.31	(US$5.45)
TP1Dflex (DurIncInPrds = 0)	US$5,129.23	TP1 (DurIncInPrds = 0)	US$5,130.54	(US$1.31)

Note. ADR = average daily revenue.

Table 5.
Average Daily Revenue and Percentage of Parties Waiting for a Table, by Model, With no Demand Timing Flexibility.

Model	Average Daily Revenue	Average Percentage of Parties Waiting for a Table Upon Arrival				
		MAT–10	MAT −5	MAT 0	MAT +5	MAT +10
TP2Dflex (extra periods = 5)	US$3,261.72	0.0035	0.0028	0.0026	0.0025	0.0031
TP1Dflex (DurIncInPrds = 2)	US$4,005.98	0.231	0.198	0.180	0.180	0.171
TP2Dflex (extra periods = 2)	US$4,013.97	0.234	0.189	0.173	0.167	0.171
TP1Dflex (DurIncInPrds = 1)	US$4,508.49	1.38	1.17	1.09	1.05	1.06
TP2Dflex (extra periods = 1)	US$4,645.86	2.49	2.00	1.77	1.71	1.71
TP1Dflex (DurIncInPrds = 0)	US$5,129.23	18.5	15.9	14.5	14.2	14.1

Note. MAT = mean arrival time relative to the designated reservation time.

Table 6.
Percentage Increase in Revenue From Higher Demand Timing Flexibility.

Model (Listed in Order of Increasing Revenue)	Demand Timing Flexibility Change (Level > Level)						
	0 > 1	1 > 2	2 > 3	3 > FF	0 > 2	0 > 3	0 > FF
TP2Dflex (extra periods = 5)	1.35%	0.84%	5.74%	NA	2.21%	8.07%	NA
TP1Dflex (DurIncInPrds = 2)	**3.38%**	**2.05%**	**10.29%**	**4.19%**	**5.50%**	**16.35%**	**21.23%**
TP2Dflex (extra periods = 2)	**3.45%**	**2.11%**	**10.92%**	**NA**	**5.64%**	**17.17%**	**NA**
TP1Dflex (DurIncInPrds = 1)	2.71%	1.28%	9.15%	3.21%	4.03%	13.55%	17.20%
TP2Dflex (extra periods = 1)	3.73%	1.50%	7.92%	NA	5.28%	13.61%	NA
TP1Dflex (DurIncInPrds = 0)	1.98%	0.16%	4.45%	NA	2.14%	6.68%	NA

Note. Results in bold are for the models near the pronounced inflection point in the service-revenue relationship.

demand shift of up to three periods and even more under full demand flexibility. We note though, that even if the parties arrive on average 10 minutes before their designated reservation time, that just over 1.2% of parties will be forced to wait for a table upon their arrival. The deterioration in customer service occurs because the restaurant is filled more fully when there is greater demand timing flexibility.

Discussion and Conclusions

The results of our study show the benefit of considering the demand timing flexibility. Compared to the assumption that demand timing is inflexible, a high level of demand timing flexibility increased revenue by as much as 21% across the models we evaluated. There are several implications from these findings. First, while modeling demand flexibility increases the size of the integer programming models and the time required to solve them, the solution time increases are modest. As such, restaurateurs should consider using demand timing flexibility.

Second, our results should allow restaurateurs to select an integer programming model and appropriate parameters that meet their requirements for maximizing revenue while maintaining high levels of customer service. The models at the pronounced inflection points in the service-level-revenue chart—TP1Dflex (DurIncInPrds = 2) and TP2Dflex (extra periods = 2)—would seem to offer the best all round performance. Interestingly, of his TP1 versions, Thompson (2015b) found the one that used the rounded-up dining duration plus one period, TP1(1) to be the most well-rounded model. In our study, because demand timing flexibility increases the number of reservations that can be accepted, it proved better to inflate the rounded-up dining duration by two periods. This highlights that parameters, and even the choice of a model, are dependent on a restaurant's specific setting.

Third, while we have demonstrated the benefits of using demand timing flexibility, the extent of demand timing flexibility in real restaurant settings has yet to be established in the literature. Demand timing flexibility is likely to be context dependent, being higher in popular restaurants, for example. Establishing the extent of demand timing flexibility is an opportunity for future research. Until such research has been conducted, and perhaps even after, restaurateurs would be well advised to go slowly: conducting some live experiments to establish the flexibility in their environment.

Fourth, customer service implications are another reason for restaurateurs to move cautiously. The higher revenue from demand timing flexibility came at the expense of lower customer service levels. However, even with full demand timing flexibility, TP1DFullFlex (round-up = 2) only required that 1.22% of parties wait for a table when the parties arrived 10 minutes early, on average. When parties arrived no more than 5 minutes early, on average, less than 1% of parties had to wait for table when they arrived.

Finally, our results demonstrate the value of a restaurant delivering the basics of great food and great service. Customers are much more likely to adapt their dining plans to the availability of the reservation slots when they know they will have a great experience. Returning to the OpenTable (2018) example described earlier, a 2-hour window before or after the desired reservation time would be equivalent to full timing flexibility in the 2-hour dining window condition we considered, and well beyond the 45-minute shifting level for the 4-hour dining window condition we considered.

In terms of limitations, we note that we did not define demand movement limits separately for earlier arrivals versus later arrivals. However, this would be a straightforward extension to the models we presented. In addition, we assumed that table sizes were proportional to the number of seats, which is not always the case in restaurants. We see no reason to expect different results for model performance or the value of demand timing flexibility, though this remains to be investigated. We believe that a worthwhile but nontrivial extension to this work would be developing an integrated optimization model for reservations and walk-in customers.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, or publication of this article.

References

Alexandrov, A., & Lariviere, M. A. (2012). Are reservations recommended? *Manufacturing and Service Operations Management*, *14*, 218-230.

Anonymous. (2017). *2017 Restaurant industry pocket factbook*. Washingon, DC: National Restaurant Association.

Bell, R., & Pliner, P. L. (2003). Time to eat: The relationship between the number of people eating and meal duration in three lunch settings. *Appetite*, *41*, 215-218.

Bertsimas, D., & Shioda, R. (2003). Restaurant revenue management. *Operations Research*, *51*, 472-486.

Dongarra, J., & Luszczek, P. (2011). LINPACK benchmark. In D. A. Padua (Ed.), *Encyclopedia of parallel computing* (pp. 1033-1036). New York NY: Springer.

Guerriero, F., Miglioni, G., & Olivito, F. (2014). Strategical and operational decisions in restaurant revenue management. *European Journal of Operational Research*, *237*, 1119-1132.

Gurobi. (2015). Houston, TX: Gurobi Optimization.

Kimes, S. E. (2008). *A consumer's view of restaurant reservation policies* (Cornell Hospitality Report, *Vol. 8*). Ithaca, NY: Cornell University.

Kimes, S. E., Chase, R. B., Choi, S., Lee, P. Y., & Ngonzi, E. N. (1998). Restaurant revenue management: Applying yield management to the restaurant industry. *Cornell Hotel and Restaurant Administration Quarterly*, *39*(3), 32-39.

Kimes, S. E., & Robson, S. K. A. (2004). The impact of restaurant table characteristics on meal duration and spending. *Cornell Hotel and Restaurant Administration Quarterly*, *45*, 333-346.

Kimes, S. E., & Thompson, G. M. (2004). Restaurant revenue management at Chevys: Determining the best table mix. *Decision Sciences*, *35*, 371-391.

Kimes, S. E., & Thompson, G. M. (2005). An evaluation of heuristic methods for determining the best table mix in full-service restaurants. *Journal of Operations Management*, *23*, 197-208.

McDougall, G. H. G., & Levesque, T. J. (1999). Waiting for service: The effectiveness of recovery strategies. *International Journal of Contemporary Hospitality Management*, *11*(1), 6-15.

OpenSolver. (2018). Available from OpenSolver.org

OpenTable. (2018). Available from OpenTable.com

Resy. (2018). Available from Resy.com

Thompson, G. M. (2010). Restaurant profitability management: The evolution of restaurant revenue management. *Cornell Hotel and Restaurant Administration Quarterly*, *51*, 308-322.

Thompson, G. M. (2011a). Cherry-picking customers by party size in restaurants. *Journal of Service Research*, *14*, 210-213.

Thompson, G. M. (2011b). Inaccuracy of the "naïve table mix" calculations. *Cornell Hotel and Restaurant Administration Quarterly*, *52*, 241-252.

Thompson, G. M. (2015a). Deciding whether to offer "Early-Bird" or "Night-Owl" specials in restaurants: A cross-functional view. *Journal of Service Research*, *18*, 498-512.

Thompson, G. M. (2015b). An evaluation of integer programming models for restaurant reservations. *Journal of Revenue and Pricing Management*, 50, 305-320

Thompson, G. M., & Kwortnik, R. J., Jr. (2008). Pooling restaurant reservations to increase service efficiency. *Journal of Service Research*, *10*, 335-346.

Author Biography

Gary M. Thompson is a tenured professor of service operations management in the School of Hotel Administration at Cornell University, where he teaches graduate and undergraduate courses in service operations management. He also serves as the Area Coordinator for Operations, Technology and Information Management in the Cornell S.C. Johnson College of Business.