How Much Is a Reduction of Your Customers' Wait Worth? An Empirical Study of the Fast-Food Drive-Thru Industry Based on Structural Estimation Methods

Gad Allon
Kellogg School of Management, Northwestern University, Evanston, Illinois 60208,
g-allon@kellogg.northwestern.edu

Awi Federgruen
Graduate School of Business, Columbia University, New York, New York 10027, af7@columbia.edu

Margaret Pierson
Harvard Business School, Harvard University, Boston, Massachusetts 02163, mpierson@hbs.edu

In many service industries, companies compete with each other on the basis of the waiting time their customers experience, along with other strategic instruments such as the price they charge for their service. The objective of this paper is to conduct an empirical study of an important industry to measure to what extent waiting time performance impacts different firms' market shares and price decisions. We report on a large-scale empirical industrial organization study in which the demand equations for fast-food drive-thru restaurants in Cook County are estimated based on so-called structural estimation methods. Our results confirm the belief expressed by industry experts, that in the fast-food drive-thru industry customers trade off price and waiting time. More interestingly, our estimates indicate that consumers attribute a very high cost to the time they spend waiting.

1. Introduction

This paper reports on an empirical study of an important industry, the drive-thru fast-food industry. We estimate a competition model, derived from an underlying mixed multinomial logit (MNML) consumer choice model, using detailed empirical data. The main goal is to measure to what extent waiting time performance, along with price levels, brand attributes, and geographical and demographic factors, impact competing firms' market shares. In the literature it is commonly assumed that customers attribute a cost rate to their waiting time that can be proxied by an earnings rate, for example, the disposable per capita income in the market (see, e.g., Mueller 1985). Our results demonstrate that this may result in questionable policy analyses because we find that customers attribute an implicit value to their wait time, which is many times the average wage in the United States. We also characterize how the market's price equilibrium responds to changes in the waiting time standards. Based on this market analysis, we show that the trend to continuously improve waiting

times and service levels can be explained on game-theoretical grounds, creating a valuable framework for future market dynamics studies in various industries. Although our empirical study is focused on the drive-thru fast-food industry, we apply a methodology based on structural estimation methods frequently used in the industrial organization literature, which, *mutatis mutandis*, can be employed to establish the impact that service attributes have on market shares in other industries.

In many service industries, companies compete with each other on the basis of the waiting time (or other service quality attributes) their customers experience, along with other strategic instruments such as their price. Executives realize that time is money for the consumer, but it is unclear how much money, how the exchange rate differs in different industries, and how it varies with other factors such as location, brand, etcetera. Often, specific waiting time standards or guarantees are advertised. For example, in 2002 Ameritrade increased its market share in the online discount brokerage market by "guaranteeing"

that equity trades take no more than 10 seconds to be executed; the guarantee is backed up with a commission waiver if the time limit is violated. This led most major online brokerage firms (E-trade, Fidelity) to offer and aggressively advertise even more ambitious waiting time standards. Various call centers promise that the customer will be helped within one hour, say, possibly by a callback. In other industries, average waiting times are monitored by independent organizations. For example, in the airline industry independent government agencies as well as Internet travel services report, on a flight-by-flight basis, the average delay and percentage of flights arriving within 15 minutes of schedule. See Allon and Federgruen (2007) for a longer list of examples.

A fundamental premise of the by now extensive theoretical literature on service competition is the belief that waiting times have a major impact on consumer choices and market shares, similar to, or perhaps even in excess of, price differentials. However, this premise has rarely been substantiated by empirical field studies. In the fast-food industry, almost all outlets are owned by independent franchisees who select their own prices. In contrast, chains set national waiting time standards by prescribing a uniform operational process to their franchisees along with specific recipes for their standard menu items. These processes include standard customer greetings, order taking, the maximum number of burgers on a grill and amount of time they may be cooking, the relationship between number of drive-thru lanes and demand volume, etcetera, all of which determine the chain's waiting time standard; see Garber (2005) and Jargon (2006). DeHoratius et al. (2010) describe how tightly the McDonald's chain standardizes and engineers the service operations process of its outlets. Because chains implement a national uniform waiting time (distribution), we were able to obtain these distributions from a national Drive-Thru Time Study Database, which we purchased from the industry organization Quick Service Restaurant (QSR).

Chains invest heavily to shave seconds off their average waiting times, clearly believing that their market shares are very sensitive to the relative waiting times experienced. Hughlett (2008) attributes the following statement to the president of one of the main technology vendors serving the fast-food industry: "There is an industry maxim that for every seven-second reduction in total service time, sales will increase by 1% over time." It is the belief expressed in this maxim that underlies the chains' continuous strategic focus on waiting time reductions in their outlets via technological and process improvements. An estimate of the expected consumer response to reductions in waiting time standards, such as that generated by our study, would be of high value to the

industry when evaluating the potential profitability of investments of this type. We therefore look to support these beliefs via an emprical study of a large fast-food drive-thru market, focusing on the following series of research questions of equal interest to the academic community and the competing chains in the (fast-food) industry.

1. Does the customer's waiting time at the fast-food drive-thru lane represent a significant determinant of consumer choices and resulting market shares, as believed by the industry and operations management literature alike?

2. When comparing the demand sensitivity to waiting time and price differentials, is the implied value of time of the same order of magnitude as the average wage or income earned per hour? If not, is it of a larger or smaller order of magnitude?

3. Can the above-stated industry maxim be substantiated by empirical estimates? In particular, when taking into account that the various outlets are likely to adjust their prices in response to reduced waiting time standards, adopting a new price equilibrium, does the maxim hold?

4. Furthermore, this maxim expresses the belief that a given waiting time reduction is equally valuable for all chains in terms of resulting increases in market shares and sales. Plausible consumer choice models may imply that these benefits, in fact, vary with the initial market shares of the chains, either in a robustly predictable way or one that depends on the specific parameter estimates. Either way, it is important to understand to what extent these benefits vary with the size of the chain and other chain attributes. In addition, do the increases in market shares accrue primarily from customers switching between chains or from the acquisition of new customers to the fast-food market?

Our empirical research thus follows a slight variant of the standard paradigm as discussed, for example, by Fisher (2007, Figure 5). Our starting point is a series of premises, maxims, and questions that arise from the theoretical operations management literature on service competition as well as practitioner discussions and surveys in the fast-food industry, reported in such outlets as the industry organization's main publication, the *QSR* magazine, the *Nation's Restaurant News*, and the general press.

As mentioned, almost all contributions to the literature on service competition have been theoretical, with numerical investigations confined to small hypothetical examples. Indeed, we believe ours to be one of the first marketwide empirical studies to complement the theoretical service competition literature. There are several reasons for the paucity of empirical studies. It is very difficult to access data regarding customer waiting times, in particular when seeking to

quantify the waiting time experience at all competing service providers. Whereas absolute waiting times at a given firm might explain the firm's demand volume in a monopoly setting, it is the relative waiting times at various competing providers that, along with the firms' other strategic choices, explain ultimate consumer choices and hence realized market shares. Similarly, it is typically very hard, if not impossible, to collect data on sales volumes or market shares of the competing outlets. Although such data are sometimes accessible for consumer products, in the service industry it is rare that sales volumes can be gathered by outsiders. Firms are reluctant to provide the information, considering it of the highest strategic value. Indeed, sales volumes were unavailable in our context. Instead, we infer them by estimating the parameters in the system of equations characterizing the unique equilibrium in a competition game resulting from a detailed consumer choice model and an outlet cost structure reflecting a broad category of queueing systems. In other words, the demand function parameters are backed out from the equilibrium conditions, with the help of the observed equilibrium. This technique has been applied in a number of economics studies, e.g., Feenstra and Levinsohn (1995) and Thomadsen (2005) but, to our knowledge, not in the operations management literature.

More specifically, we accommodate the absence of demand data with three assumptions: (1) Consumers attribute a utility level to each potential outlet that depends stochastically on price, waiting time, the distance to the outlet, and various chain characteristics. Similarly, consumers assign a utility level to the no-purchase option, which depends stochastically on the consumer's gender, race, age bracket, and occupational status. (2) Outlets encounter a cost structure that is affine in the sales volume, with random noise terms for the marginal costs; this cost structure applies to many queueing models used to describe the service process, such as $M/M/1$ systems or open Jackson networks. (3) Outlets adopt a pure Nash equilibrium in the price competition model that results from the above consumer choice model and the outlets' cost structure.

The first assumption is used to derive the relationships between prices, service levels, and sales quantities. Based on the second and third assumptions, these relationships are subsequently used to derive the firms' Nash equilibrium conditions to jointly estimate the parameters of the indirect utility functions of the consumers as well as the parameters of the outlets' cost structure. Our estimation method is a generalized method of moments (GMM) technique, as opposed to more standard maximum-likelihood estimators for systems of nonlinear equations, for reasons explained in §5.

In summary, the main contribution of this paper is that to our knowledge it is one of the first to estimate, for the benefit of market observers and the firms alike, how sales volumes for a service organization depend on the prices and waiting times of all competing providers within a given region, their location, as well as other attributes (e.g., brand-specific characteristics). In particular, we conclude that consumers attribute a value to their waiting time that is many times the average wage level. We use counterfactual studies to confirm that a seven-second reduction by a single chain results, on "average," in a 1% market share increase for that chain. However, for a large chain like McDonald's, the increase is more than 3%, showing that the industry's "seven second rule" needs to be qualified. The increased market share results primarily from the acquisition of new customers who were previously opting for the outside good as opposed to cusomers switching between chains. Our model explains the continuing trend of all chains investing heavily to reduce their waiting time standards. We show, in addition, that neglecting to include any waiting time measure in the consumer choice model results in significantly overestimated price sensitivities. This validates our belief that overlooking service as a competitive instrument in the model specification results in distorted managerial insights. Not accounting for the waiting time as an attribute also distorts the estimated value of the no-purchase option, as well as the importance of the number of chain outlets, as a proxy for the consumers' perceived quality of the chain.

The remainder of this paper is organized as follows: Section 2 provides a review of the relevant literature. Section 3 develops our consumer choice and competition model. Section 4 describes the many data sources employed and the approach we adopted to collect the data. Section 5 is devoted to a description of the GMM estimation technique as applied to our model. Section 6 describes the estimation results and counterfactual studies. Section 7 completes the paper with a discussion of possible extensions.

2. Literature Review

The literature on competition in service industries dates back to the late 1970s. Luski (1976) and Levhari and Luski (1978) were the first to model competition between service providers. The latter paper addresses a *duopoly* where each of the firms acts as an $M/M/1$ system, with exogenous and identical service rates. In this model, customers select their service provider strictly on the basis of the *full price*, defined as the direct price plus the expected steady-state waiting time multiplied with the waiting time cost rate. The question of whether a price equilibrium exists in

this model remained an open question until it was recently resolved in the affirmative by Chen and Wan (2003), albeit for the basic model with a uniform cost rate. These authors show, however, that the Nash equilibrium may fail to be unique.

Cachon and Harker (2002) and So (2000) analyzed the first models in which customers consider criteria beyond the lowest full price when choosing a service provider (e.g., quality). Both confined themselves, again, to $M/M/1$ service providers. Allon and Federgruen (2007, 2008) treat the price and waiting time standard as completely independent firm attributes that different customers may trade off in different ways. See those, as well as Hassin and Haviv (2003), for additional references of service competition models.

Many service processes are provided via call centers. Here, customers are known to be very sensitive to their waiting times, which is why such centers are designed and staffed to meet specific service-level agreements (SLAs); see Hasija et al. (2008) for a recent survey of such agreements. However, virtually all planning models in the vast literature related to call centers assume that demand processes are exogenous inputs, or, at best, dependent on service charges. We refer to Gans et al. (2003) for an excellent tutorial on call center management. When describing future challenges in this area, the authors emphasize "a better understanding of customer behavior" (§7.3) and the need to model and estimate "multiple levels of equilibria." Beyond these levels, we suggest the desirability of models incorporating the competitive effect of service levels provided by the call centers of competing service providers.

The above-reviewed literature is based on the observation that firms compete along the service-level dimension as well as anecdotal and empirical evidence that customers value waiting time when making decisions regarding their preferred service provider. One example in the fast-food industry is Davis and Vollmann (1990), who examined consumer choice criteria with a sample of 723 customers who were asked to rank their satisfaction with various aspects of the delivery process. The authors established that the satisfaction scores were highly correlated with the experienced waiting time. Time of day, store location, and whether the customer was at work or school were important factors determining the strength of the waiting time sensitivity. Day of week and participation activities other than work around the meal (e.g., shopping, visiting friends) were not significant.

This empirical study complements the earlier quoted plethora of trade literature documenting the centrality of waiting times in this industry. Our study complements the theoretical literature on competition models by estimating the parameters used and

assumed by these models. There have been very few other works that attempt to estimate these parameters. The two empirical studies investigating questions closest to our own are Deacon and Sonstelie (1985) and Png and Reitman (1994). The former appears to have been the first to estimate the impact price differentials and average waiting times have on sales volumes; however, the setting is one where prices are exogenously determined by government price controls, avoiding the endogeneity challenge inherent in most studies, including our own. The selected estimation method is based on a probit model, applicable only in the case of two firms. The model does not apply to settings with price-selecting firms or those where customer choices depend on factors other than the full price. Png and Reitman (1994) describe the peak-hour sales in a market of 1,501 gas stations in four Massachusetts counties via a system of demand equations. These equations are not derived from an underlying consumer choice model as, for example, in Deacon and Sonstelie (1985) or our paper. In the absence of actual observations of the waiting times or the peak-hour sales volumes, the authors specify the logarithm of a firm's (peak-hour) sales volume as a linear function of the logarithm of the firm's own price, that of the average of the prices of the "nearby" stations, a vector of station attributes, and a proxy for the average waiting time. The latter is postulated as the ratio of the peak-hour sales volume and a predetermined power of the number of pumps, whereas the peak-hour sales volume is assumed to be given by the aggregate weekly sales divided by a given power of the number of operating hours. The coefficients in this model are assumed to be homogeneous constants that are estimated via a least-squares regression method. The authors address the problem of the explanatory prices and capacity variables being endogenous to the system by the use of a two-stage least-squares method, invoking instrumental variables claimed to be uncorrelated with the error terms.

The transportation research literature has often specified the demand for alternative transportation modes as arising from a mixed multinomial logit model with prices and travel times as explanatory variables, similar to our paper. However, the maximum-likelihood estimations typically employed are challenged by the abovementioned endogeneity problems; see Hess et al. (2005) for a recent example, estimating the implicit cost associated with travel time to be in excess of $100/hour. Finally, an earlier economics paper by De Vany et al. (1983) estimated the effect waiting times have on patient volumes in dentist offices; ignoring the impact of competition and employing OLS, the authors obtain a statistically significant positive value for the waiting time sensitivity, perhaps because in their setting demand is relatively

inelastic with respect to waiting times while capacity is inflexible. In contrast, we estimate the impact of waiting times, prices, geographic dispersion, chain attributes, and demographic factors on demand. Our approach follows the work of Bresnahan (1987), Berry (1994), and Berry et al. (1995). In what came to be known as the Berry–Levinsohn–Pakes (BLP) method, these authors demonstrate how to estimate consumer choice models and cost structures in oligopolistic markets with differentiated goods using aggregate consumer-level data and structural models of competition. The general approach posits a distribution of consumer preferences for the competing goods based on their attributes. The preferences are aggregated into a market-level demand system that when combined with assumptions on cost and price-setting behavior allows one to estimate the parameters.

In the above papers, market shares are observed. Feenstra and Levinsohn (1995) were the first to demonstrate how this estimation framework can be used in the absence of quantity data. As mentioned in the introduction, we face the same challenge because in the fast-food industry, sales data are not reported and are treated as strategic and proprietary information.

More recent work by Davis (2006) and Thomadsen (2005) incorporated geography in the BLP framework. Thomadsen (2005) studies the impact of ownership structure on prices in the fast-food industry. The author uses this method to establish that the impact of mergers in such an industry can be large, but the impact of mergers decreases as the merging outlets are further apart. Our consumer choice model adds the waiting time measure to the set of outlet and chain-dependent explanatory variables employed in Thomadsen (2005), but does not include ownership structure. For reasons explained in the model section, we incorporate other chain attributes that act as indicators of perceived quality, instead of the chain dummy variables employed in Thomadsen (2005).

Our study is also related to the recent empirical literature in operations management. To our knowledge, most of this literature focuses on consumer products rather than services. See Olivares and Cachon (2009) and Musalem et al. (2010) for surveys of this literature. Notable exceptions include Olivares et al. (2008) and KC and Terwiesch (2009, 2011), which focus on the health-care industry.

3.   The Model

In this section we develop the competition model representing the competitive interdependencies and interactions among the outlets in our geographic region. The model combines two submodels: (a) a *consumer choice* model that determines how many of the region's residents and commuters choose, for any given lunch or dinner meal, to go to a fast-food establishment and, among those, how many select a specific outlet; and (b) a model to represent the *variable cost* structure of the different outlets as a function of their sales volume and service level (i.e., its waiting time standard). Combining the two submodels permits us to derive the outlets' profit functions. As explained, in the fast-food industry, waiting time standards are selected by the chains. However, price decisions are relegated to the independent outlets, if for no other reason than to avoid illegal forms of price fixing. As franchising became popular in the sixties, the U.S. courts began to limit the types of pricing restrictions that chains can impose on their franchises. Only maximum retail prices have become legal, under certain conditions, based on the Supreme Court ruling in *State Oil Co. v. Khan*.[1] (In the prior 30 years, even maximum price levels had been illegal; see *Albrecht v. Herald Co.*[2])

When collecting data, we called the *chains* for price recommendations that they may give to their franchisees. Consistent with the Supreme Court rulings, we were told that the practice of *suggesting* prices to the outlets is illegal. Indeed, we have observed significant price differences among outlets of the same chain, see Table 2 in §4. Thus, waiting time standards are selected centrally by the chains, but prices are chosen by the individual outlets. We can therefore assume that the prices observed in the market represent the equilibrium in a price competition model, under given waiting time standards specified by the chains operating in the selected geographical region. We show that this price equilibrium model has an equilibrium that is a solution of a nonlinear system of equations. It is this system of equations that permits us to estimate the parameters that describe the consumer choice model and associated demand functions, as well as the parameters in the cost structure.

3.1.   The Consumer Choice Model

Demand for fast-food meals at each outlet is specified by a discrete choice model. Consumers choose either to purchase a specific lunch or dinner meal from one of the fast-food outlets or to consume an outside good. Consumers assign a utility value to each outlet, as well as to the no-purchase option, specified as a linear function of the price, waiting time, distance, chain identity of the outlet, and various demographic factors including the consumer's gender, race, age bracket, and occupational status. Each of these utility equations contains an additional random noise term. It is natural to assume that customers make their choices in two stages: (i) they first decide whether to dine at a fast-food outlet as opposed to

---

[1] *State Oil Co. v. Khan*, 522 U.S. 3 (1997).

[2] *Albrecht v. Herald Co.*, 390 U.S. 145 (1968).

alternatives, such as eating at home or a different type of restaurant; and (ii) assuming the first question is answered in the affirmative, which of the various outlets in the region to patronize. We model the two-stage choice process by assuming that the (potential) customer attributes a utility value to the no-purchase option that depends on her demographic attributes. The customer also assigns a utility value to each of the outlets in the region that depends on attributes of both the outlet and the chain to which it belongs. The customer purchases a meal at one of the fast-food outlets if and only if the highest of the outlets' utility values is in excess of that of the no-purchase option; in this case the meal is consumed at the outlet with the highest utility value. Formally, the conditional indirect utility of consumer i from fast-food outlet j is specified as follows:

$$U_{i,j} = \beta + X'_{k(j)}\zeta - \delta D_{ij} - \gamma P_j - \alpha W_{k(j)} + \eta_{ij}, \quad (1)$$

where $k(j)$ denotes the chain k to which outlet j belongs; $X_{k(j)}$ is a column vector of observed properties of the chain to which outlet j belongs; D_{ij} is the distance between consumer i and outlet j; P_j is the price of a (standard) meal at outlet j; $W_{k(j)}$ is the waiting time standard (equal to average steady-state waiting time in system) of chain $k(j)$ associated with outlet j; η_{ij} is the portion of the utility of individual i at outlet j, which is unobserved by the modeler; and $(\alpha, \beta, \gamma, \delta, \zeta)$ represents a parameter string with ζ an array of the same dimension as X.

Our estimation of waiting time sensitivity is based on three assumptions: (i) consumers make purchasing decisions based on the steady-state waiting time distribution at an outlet, not on the prevailing queue length (the only varying observable characteristic) at the time of arrival; (ii) consumers characterize the steady-state waiting time distribution by its average; and (iii) all outlets belonging to the same chain share the same waiting time distribution. The first assumption is based on our understanding that in most cases consumers make their selection before traveling to any specific outlet based on the "average" experience. The second assumption is not inherent to our approach and could easily be replaced by other characteristics of the steady-state waiting time distribution, such as the 95th percentile. As for the final assumption, we explained earlier that in the fast-food industry the chains select and announce to their franchisees a common waiting time standard for all of their outlets, implemented with tight process prescription and control. Indeed, chains achieve remarkably uniform average waiting times at their franchises. Jargon (2006) reported that the average waiting time at Burger King's 6,900 domestic franchises covers a very narrow range from 165 to 170 seconds; the average waiting time was reduced

by 22 seconds in one year. Industry trade organizations such as QSR publicize yearly surveys of the average waiting time experienced at the various fast-food chains. Outlet-specific samples in our waiting time data set are too small to make our own empirical verifications of assumption (iii). Furthermore, although there are empirical papers that have individual waiting time observations for a single outlet or chain in a service industry, such a data set is extremely difficult to obtain for industrywide studies.

Because all outlets belonging to the same chain share the same waiting time standard, it is important to include in the individual utility functions (1) any other observable chainwide attributes that (i) are correlated with the waiting time standard and (ii) may plausibly serve as a quality indicator for the chain. The only such attributes we were able to identify are the density of the chain network, as measured by the number of outlets, in the county and the intensity of the chain's advertising efforts, as quantified by its aggregate national advertising spending. We do not use chain identity indicator variables, as is frequently done, because its inclusion among the explanatory variables in (1) results in an identification problem. Normalizing the coefficient of one of the chain indicator variables removes collinearity in the utility function; however, market shares remain invariant to a common additive shift in the coefficients of the indicator variables and that of the waiting time standard.

A recent QSR-commissioned study (Magid Associates 2010) reports on a survey among 1,120 drive-thru customers, in which each respondent was asked to list which of 10 attributes makes the drive-thru experience the best. The most frequently cited attribute was "speed—wait time of drive-thru service" (22% of respondents), followed by "price" and "order accuracy" (12% each). Location was listed almost as frequently as "price" (and order accuracy), i.e., by 11% of respondents. Our consumer choice model represents all of these attributes with the exception of "order accuracy;" the latter should be included in future studies, in particular because it may shed light on an optimal balance between "speed of service" and "order accuracy."

We assume that for every outlet j, the random components of $\{\eta_{ij}\}$ represent nonsystematic unobservable variations in the perceived utility of the outlet among potential customers of the same demographic type residing or working in the same location. We therefore assume that the $\{\eta_{ij}\}$ variables are independent and identically distributed (i.i.d.) random utility models of type (1) often contain an additional outlet-specific component ξ_j, $j = 1, \ldots, N$, to address systematic attributes of the firm (outlet), known to the firms and customers, but not to the modeler. As argued in Thomadsen (2005), in the case of the fast-food drive-thru industry, this term may be omitted

because other than through price and location, different outlets belonging to the same chain offer close to identical attributes, whereas the chains create virtually identical "experiences" at all their locations. At the same time, as explained above, all relevant chain-specific attributes are captured by the chain variables X in the first term in (1), along with the waiting time standard W. The indirect utility associated with the no-purchase option is given by

$$U_{i,0} = \beta_0 + M_i \pi + \eta_{i,0}. \quad (2)$$

Here, M_i is a row vector specifying the consumer's age, gender, race, and whether they are making the decision as a commuter or resident (i.e., people are allowed to have a different preference for the outside good when they are at work versus at home). If the age distribution is characterized by A age classes, the M_j vector is a binary vector of dimension $(A + 2)$: for $l = 1, \ldots, A-1$, $M_{il} = 1$ if consumer i belongs to the lth age bracket and 0 otherwise; similarly, for $l = A$, $A+1$, and $A+2$, $M_{il} = 1(0)$ if the consumer is female (male), African American (white), and a resident (worker), respectively. Terms β_0 and π represent another set of parameters to be estimated, and η_{i0} denotes the unknown portion of the utility of individual i for the nonpurchase option. Once again, the random components $\{\eta_{i0}\}$ are i.i.d.

We consider a limited number of age brackets. Therefore, there is a finite list of $\{1, \ldots, M\}$ of consumer types, combining age, gender, race, and occupational status. In view of the importance of the distances between the consumer and the various outlets, we partition our geographic region into a grid of very small subareas, $\mathcal{B} = \{1, \ldots, B\}$, and assume all consumers residing in a subarea are located at the subarea's centroid. (In our study, we use tracts, as defined by the U.S. Census, with an average area of 1.2 square miles in Cook county.) Thus, all potential consumers residing in a given subarea, $b \in \mathcal{B}$, and belonging to a given demographic group, $m \in M$, share the same *mean* utility value for all outlets and the no-purchase option.

Assuming that the distributions of the random noise terms, $\{\eta_{ij}: j = 0, \ldots, J\}$, are Gumbel (or doubly exponential) with common scale parameter μ, and assuming that every consumer selects the alternative with the highest utility value, this gives rise to the following multinomial logit model in which each outlet's market share for each tract and demographic group is given by the following expression:

$$S_{j,b,m}(P, W, X \mid \beta, \zeta, \delta, \gamma, \alpha, \pi)$$
$$= \frac{e^{(\beta + X'_{k(j)}\zeta - D_{bj}\delta - P_j\gamma - W_{k(j)}\alpha)/\mu}}{e^{(\pi M + \beta_0)/\mu} + \sum_{t=1}^{J} e^{(\beta + X'_{k(t)}\zeta - D_{bt}\delta - P_t\gamma - W_{k(t)}\alpha)/\mu}},$$
$$j = 1, \ldots, J; \ b = 1, \ldots, B; \ m = 1, \ldots, M. \quad (3)$$

Without loss of generality, we express the utility levels in units such that the scale parameter $\mu = 1$. Also, the consumer choices only depend on the *relative* ranking of the utility values for the different outlets and the no-purchase option; they are therefore invariant to a common additive shift. This permits us to normalize the intercept β in the utility function (1) to $\beta = 0$.

Multiplying the market shares with $h(b, m)$, the number of consumers of demographic group m residing in or commuting to geographic region b allows us to specify expected aggregate sales in an outlet as a function of the various parameters $\theta \equiv \{\zeta, \delta, \gamma, \alpha, \pi\}$ in the utility equations:

$$Q_j(P, W, X \mid \zeta, \delta, \gamma, \pi)$$
$$= \sum_b \sum_m h(b, m) S_{j,b,m}(P, W, X \mid \zeta, \delta, \gamma, \pi,). \quad (4)$$

3.2. The Outlets' Cost Structure

When assessing the impact of operational measures, it is important to specify a cost structure which is rigorously substantiated by an adequate operational model. We have selected a structure, in which an outlet's costs, expressed as a function of its expected sales volume, is affine with an intercept that is proportional with the reciprocal of the waiting time standard:

$$C_j(Q_j) = \bar{c}_j Q_j + \bar{d}_j/W_{k(j)}$$
$$= (c_{k(j)} + \epsilon_j)Q_j + [d_{k(j)} + u_j]/W_{k(j)},$$
$$j = 1, \ldots, J. \quad (5)$$

Here, for every outlet $j = 1, \ldots, J$ and chain $k = 1, \ldots, K$:

$J_k =$ the set of outlets belonging to chain k, i.e., $J_k = \{j: k(j) = k\}$;

$c_{k(j)} =$ the average variable food, labor, and equipment cost rate per customer for an outlet of chain k;

$d_{k(j)} =$ the average variable capacity cost rate for an outlet of chain k;

$\epsilon_j =$ a noise term, denoting the difference between outlet j's variable cost rate $\bar{c}_j$ and the norm or average for this chain $c_{k(j)}$;

$u_j =$ a noise term, denoting the difference between outlet j's variable capacity cost rate $\bar{d}_j$ and the norm or average for this chain $d_{k(j)}$.

Each outlet's marginal cost rate, as well as the capacity cost rate, is equal to a common chain-specific cost plus a zero-mean, unobserved outlet-specific component. This specification is supported by the franchisers' effort to create a uniform customer experience across their outlets via standardization of the equipment, as well as the preparation process and food components used at each of its outlets. The unobserved shock to the cost rate comes from outlet-specific conditions (e.g., deficiencies in labor productivity, management

efficiency, or smaller kitchens creating crowding and reduced efficiency).

The affine cost structure in (5) arises in several queueing models that may describe the service process of an outlet. For example, the structure in (5) arises in an $M/M/1$ system, where the waiting time standard W denotes the expected total sojourn time in the drive-thru queue and the variable capacity cost is assumed to be proportional with the service rate. More realistically, a fast-food service process could be represented as a Jackson (queueing) network. A food order may travel along a path of service stages, from order taking to the cooking of the hamburgers, assembly of the cooked burgers with the side dish and required drink, and back to the drive-thru counter. Allon and Federgruen (2008) have shown that the cost structure in (5) applies to a general Jackson network, assuming that the variable capacity costs are proportional with the service rates installed at the various nodes of the network. Alternatively, the service process may be best described as a $GI/GI/s$ system, with an arbitrary renewal arrival process, arbitrary service time distribution, and a team of s parallel servers. If the consumer is particularly focused on the *delay* experienced in the drive-thru queue and if W denotes a given fractile of the delay distribution, then the cost structure in (5) arises as a close approximation; see Allon and Federgruen (2008). This identity is in fact exact, rather than an asymptotically correct approximation, when the service time distribution is exponential, i.e., in the case of a $GI/M/s$ system.

We refer to Allon and Federgruen (2008) for additional queueing models resulting in affine cost structures of type (5). These authors also show that an even larger set of queueing models give rise to a more complex family of cost functions. Our estimation method, which fits the model parameters to the first-order conditions (FOC) of the underlying competition model, can be adapted to this more general cost structure; see §3.3 for more discussion. However, Allon and Gurvich (2010) show that approximating a more complex capacity function by an affine function results in only minor discrepancies in the price equilibrium. Thus, disregarding higher-order terms does not significantly alter the outcomes of the market.

3.3. The Price Competition Model

We are now ready to analyze the price competition model that arises when all waiting time standards have been specified. We assume that every outlet is independently owned. However, our methodology is readily adapted if various outlets are jointly managed by the same franchisee; see below. In view of (5),

$$\pi_j(P, W, X, \theta) = (P_j - \bar{c}_j)Q_j(P, W, X \mid \theta) - \bar{d}_j/W_{k(j)},$$
$$j = 1, \ldots, J \quad (6)$$

denotes firms j's profit level as a function of all prices charged by the various outlets. Each firm j selects its price within a given range $[\bar{c}_j, p_{j,\max}]$. It is a long-standing conjecture that a price competition model with a mixed multinomial logit demand function and an affine cost structure has a *unique* interior point equilibrium that is the unique solution of the system of equations given by the FOC:

$$Q_j(P, W, X \mid \theta) + (P_j - c_{k(j)} - \epsilon_j)\frac{\partial Q_j(P, W, X \mid \theta)}{\partial P_j} = 0,$$
$$j = 1, \ldots, J. \quad (7)$$

This conjecture underlies almost all structural estimation methods in models with demand equations of this type. Indeed, the essence of these estimation methods is to find parameter combinations under which the FOC equations (7) are satisfied as closely as possible since the competing firms are assumed to have adopted the observed price vector as a Nash equilibrium.

Unfortunately, little was known about whether or when the above conjecture holds, see e.g., Berry et al. (1995) Indeed, Allon et al. (2009) show that an equilibrium may fail to exist in the general model without *any* parameter restrictions. A Nash equilibrium does exist and the set of equilibria corresponds with the set of solutions to (7), provided one can ensure that no single firm attains an excessively large share of the market when pricing at a specific level, which under the condition, is shown to be an upper bound for a rational price choice. More specifically, the authors introduce the following parameterized condition:

$C(\mu)$. Each firm j captures, in each market segment—i.e., each tract/demographic group combination (b, m)—less than a fraction of the market when pricing at the level $\bar{p}_j = \bar{c}_j + 1/(1 - \mu)\gamma$, $j = 1, \ldots, N$. This condition is easily satisfied by checking for every firm j and market segment (b, m) that

$$\frac{e^{[X'_{k(j)}\zeta - D_{b,j}\delta - \bar{p}_j\gamma - W_{k(j)}\alpha]}}{e^{(\pi M - \beta_0)} + e^{[X'_{k(j)}\zeta - D_{b,j}\delta - \bar{p}_j\gamma - W_{k(j)}\alpha]} + \sum_{t \neq j} e^{[X'_{k(t)}\zeta - D_{b,t}\delta - p_t^{\max}\gamma - W_{k(t)}\alpha]}} \leq \mu.$$
$$(8)$$

A firm's market share is, of course, maximized when its competitors adopt their maximum price levels. (See Allon et al. 2009 for sufficient conditions of $C(\mu)$ that are independent of the choice of the $\{p_j^{\max}\}$ values.) Under condition $C(\mu)$, Allon et al. (2009, Lemma 4.1) show, in fact, that for all $j = 1, \ldots, N$, the price level $\bar{p}_j = \bar{c}_j + 1/(1 - \mu)\gamma$ arises as an upper bound for firm j's price level. Of particular importance are conditions $C(1/2)$ and $C(1/3)$, which ensure that no outlet captures more than 50% or 33% of the market, respectively (under the abovementioned price levels). Conditions $C(1/2)$ and $C(1/3)$ are easily satisfied in most industries and the fast-food industry in particular. (See §6 for a verification.) Theorems 4.2

and 4.3 in Allon et al. (2009), applied to our model, imply the following:

THEOREM 3.1. (a) *Under condition* $C(1/2)$, *the price competition model has an equilibrium that is an interior point of the price cube* $X_{j=1}^{N}[\bar{c}_j, \bar{c}_j + 2/\gamma]$.

(b) *Under condition* $C(1/2)$, *every Nash equilibrium is a solution to the FOC* (7), *and, vice versa, every solution to* (7) *is a Nash equilibrium.*

(c) *Under condition* $C(1/3)$, *the price competition model has a unique equilibrium that is an interior point of the price cube* $X_{j=1}^{N}[\bar{c}_j, \bar{c}_j + 1.5/\gamma]$; *this equilibrium is a solution to the FOC* (7).

Given the affine cost structure (5) as substantiated in §3.2 on the basis of underlying queueing models, the waiting time measures impact only via the demand model, i.e., via the marginal price sensitivities $\{\partial Q_j(P, W, X \mid \theta) \partial p_j: j = 1, \ldots, J\}$; see (3), (4). As discussed in §3.2 under more general queueing models, the marginal costs become a function of the waiting time measures as well, in which case the vector W impacts the structure of the FOC equations both via the demand and the supply models. See also §6.

4. Data

We have studied the hamburger drive-thru fast-food industry in Cook County, Illinois. We chose this industry both because of the availability of data and because this is an industry that has historically placed a premium on competing via its service levels. The *QSR* 2007 Drive-Thru Time Study notes that in 2007 all quick-service chains made major efforts to improve speed of service in their drive-thrus (see Nuckolls 2007). Examples of new technology improving speed of service include timer systems that allow in-store managers, as well as regional and national offices, to monitor waiting times at outlets and the outsourcing of drive-thru order taking. The 2008 *QSR* Drive-Thru Study reports that this trend is continuing, with the fastest chain, Wendy's, shaving off an additional seven seconds from the average waiting time in the previous year. There is a plethora of anecdotal evidence that the industry is reacting to consumer expectations regarding waiting times. For example, the same 2007 *QSR* Drive-Thru Study reported that 70% of surveyed customers said speed is an important factor in the drive-thru experience.

We believe that for our purposes Cook County is representative of all urban/suburban counties in the country. The propensity to consume hamburger fast-food meals as opposed to alternatives may differ in different parts of the country. However, we see no reason why within urban/suburban areas, the relative trade-offs between price, waiting times, geography, and other chain attributes among those interested in a fast-food meal would vary significantly.

We use as our data set all fast-food outlets belonging to chains selling hamburgers and with a presence of more than five outlets in the county. We consider only outlets with drive-thru windows because outlets without drive-thru windows tend to be located in places such as malls and airports where consumers are facing a different set of considerations. This results in a total of 388 outlets belonging to McDonald's (173), Burger King (92), Wendy's (62), White Castle (42), Dairy Queen (10), and Steak 'n Shake (9).

Our consumer choice model does not differentiate among the various items on the outlets' menus. Our model choice is based on the assumption that consumers, when trading off different outlets (as well as the no purchase option), consider a *general* price assessment about each restaurant rather than a complete comparison of all fully itemized menus, information they are unlikely to possess, let alone be able to aggregate in a comprehensive trade-off among alternative outlets. We have demonstrated that this trade-off requires the consideration of waiting time, geography, and chain attributes along with the general price level. As a proxy for the general price level of a hamburger drive-thru restaurant, we have computed the price of a "standard meal" consisting of the franchise's signature burger, a small fries order, and a small soft drink. (We gathered prices by calling each location.) The type of burger selected was standardized by weight, and in the case of White Castle, which sells small burgers, we use the price for four sliders. As noted in the introduction to §3, we have observed very significant price differences among outlets belonging to the same chain, with the most expensive McDonald's or Burger King outlet being about 50% more expensive than the cheapest outlet of that chain in the county; see Table 2.

Recall that in the absence of data on joint ownership among franchises, we assume each outlet is owned independently, or at least operating as an independent profit center. In the state of Texas, there is a limited amount of joint ownership among franchisees (see Kalnins and Lafontaine 2004). Also, the larger chains allow multiple-unit ownership only via purchases of individual units and only under certain conditions. The only exception among the six chains in our study is White Castle, all of whose outlets are owned by the chain. Because all of the outlets sell similar products, i.e., hamburgers, and use the same means of service, i.e., drive-thru facilities, we assume that all of these outlets compete with each other as alternative producers in the same market.

We use two chain attributes in addition to the chains' waiting time standard as potential indicators of the consumer's perception of chain quality: (1) the density of the chain network measured by the number of outlets in Cook County; and (2) the "intensity" of the chain's advertising efforts, as measured

Table 1   Average Waiting Time as Determined from the 2005 *QSR* Drive-Thru Study

Chain	Mean wait (sec.)	No. of outlets	National advertising spend ('05 MM$)
Wendy's	173.34	62	360
Burger King	192.29	92	265
McDonald's	224.27	173	638
Dairy Queen	231.85	10	56
Steak 'n Shake	264.30	9	12.5
White Castle	269.45	42	12.5

by the national advertising spend. The advertising spend was taken from the Ad$pender database using the "Total Ad Spend" figure for the "General Promotion" category for each chain in 2005. All media outlets were included (e.g., television, U.S. Internet, radio, and print). As discussed, all chains select and strive for a common waiting time standard among all of their outlets. In addition, customers often frequent more than a single outlet of a chain and expect to experience a similar service level, irrespective of the specific outlet they visit. We have selected the average steady-state waiting time, defined as the time spent in the drive-thru queue plus the service time, as the waiting time standard used in the consumer choice model of §3.1. To arrive at the average waiting time standards for the different chains, we have employed the 2005 Drive-Thru Time Study Database, which we purchased from QSR. The database contains, for a national sample of outlets, two random observations at lunch and at dinner time. We obtained each chain's average waiting time by averaging the recorded observations over all outlets that belong to the relevant chains, nationwide. These national average waiting times vary significantly across chains, with the worst performer being close to twice as slow as the best performer; see Table 1. The chain wide waiting time standards of the six chains in our study have a mean of 225.92 seconds, a standard deviation of 38.21, and a range of [173.34, 269.45]. Using a two-sample *t*-test assuming unequal variances on all of the national waiting time observations, we have verified that waiting time observations for the six different chains were indeed drawn from different distributions with the exception of two pairs: McDonald's and Dairy Queen, and White Castle and Steak 'n Shake. (Note from Table 1 that the mean waiting times are nearly identical within each of these two pairs as well.) This confirms that different chains offer systematically different waiting time experiences to the consumer. The results of this analysis can be seen in Table A.1 in Appendix A.

Demographic and geographic information was gathered with a very fine granularity at the so-called tract level. Tracts are geographic areas defined by the U.S. Census Bureau to contain 2,500 to 8,000 people.

Cook County consists of 1,343 tracts with an average area of only 1.2 square miles. In urban areas a tract corresponds with a few city blocks. The next-smallest geographic area recognized by the U.S. Census, the so-called block groups, are so small that some demographic data, such as race, cannot be reported without revealing the exact household being discussed, and hence are not available to the public. We have considered the following age brackets: 0–9, 10–19, 20–39, 40–59, and 60+. We considered African American and white consumers only because these are the racial groups for which we had the necessary data for employing the macro moments discussed in §6. As mentioned in §3, consumers are also differentiated based on whether they are at work or home. As far as the residents in a tract are concerned, we collected the number of people of each age bracket, race, and gender combination from the 2000 U.S. Census data. As to the population working in each of the tracts, the Bureau of Transportation Statistics reports the number of people commuting between every tract pair. We aggregated the flow of workers into each tract in Cook County from any originating tract (whether or not the originating tract was within Cook County). Unlike the U.S. Census data, the Bureau of Transportation Statistics data are not broken down by age, gender, and race combinations, so we estimated the population size for each triplet combination by assuming that the three demographic attributes are independent. If a person lives and works in Cook County they are counted as two consumers. We do this because such consumers have the potential to consume one meal (e.g., lunch) while at work and another meal (e.g., dinner) while at home. Distinguishing between commuters and residents, two genders, and two racial groups, as well as among five age brackets, we have thus divided the population into 40 different demographic groups. The distance from the consumer to each outlet is calculated as the distance between the restaurant and the centroid of the tract in which the consumer is located. To compute these distances, we employed the ArcView Geographic Information System modeling and mapping software.

In addition to the independent variables, we collected data for the so-called instruments used in the estimation method. As discussed in the next section, these are outlet-specific variables that we argue are correlated with one or more of the independent variables, but not with the noise terms $\{\epsilon_j: j = 1, \ldots, J\}$ in the cost rates, i.e., the outlet-specific shock on chain-wide marginal cost. Following the recommendation in Thomadsen (2005), we have selected the following instrumental variables: V_{1j} is the distance from outlet j to the nearest outlet, V_{2j} is the number of outlets within two miles of outlet j, V_{3j} is the population density in the tract to which outlet j belongs, and V_{4j} is

Table 2   Summary Statistics for Outlet-Specific Data

Variable name	Mean	Standard deviation	Min	Max
McDonald's price ($)	4.96	0.25	4.20	6.09
Burger King price ($)	4.85	0.28	3.63	5.39
Wendy's price ($)	4.75	0.20	4.27	5.24
White Castle price ($)	4.46	0.09	4.23	4.78
Dairy Queen price ($)	5.66	0.26	5.07	6.07
Steak n' Shake price ($)	4.99	0.36	4.67	5.84
Distance to nearest outlet (mi)	0.55	0.48	0.00	2.52
No. outlets within 2 mi	5.93	2.54	1	14
Population density (100 K/sq mi)	0.09	0.09	1.71E–04	0.80
Worker density (100 K/sq mi)	0.04	0.05	1.33E–03	0.36

the worker density in this tract. Table 2 shows summary statistics for these instruments as well as the price variables.

5. Estimation

As mentioned in the introduction, the major hurdle when estimating the parameters of the demand functions and the firms' cost structure is the lack of available demand data. As explained, this challenge is not unique to the fast-food industry, but presents itself in almost all service industries. Because of the unavailability of sales data, we employ a technique that estimates the parameters on the basis of the system of FOC equations (7), the solutions of which, by Theorem 3.1, coincide with the Nash equilibria of the price competition model.

The equilibrium conditions (7) represent a system of equations that involve only the *observed* price vector P, waiting time standards W, outlet attribute matrix X, and distances $\{D_{j,b}\ j = 1, \ldots, J,\ b = 1, \ldots, B\}$, as well as the *unknown* parameter string and cost-rate residuals. (In particular, the system of equations does not involve the unobservable sales volumes.) The system of Equation (7) allows us to determine the cost-rate residual as closed-form functions of the observed explanatory variables and unknown parameters. It is easily verified that

$$
\frac{\partial Q_j(P, W, X \mid \theta)}{\partial P_j}
$$

$$
= -\gamma \sum_{b=1}^{B} \sum_{m=1}^{M} h(b, m)\left(1 - \frac{S_{j, b, m}(P, W, X \mid \theta)}{h(b, m)}\right)
$$

$$
\cdot S_{j, b, m}(P, W, X \mid \theta). \tag{9}
$$

In matrix notation, the equilibrium conditions (7) can be stated as

$$
Q(P, X, W) + \Omega(P - \bar{c}) = 0, \tag{10}
$$

where Ω is a diagonal $J \times J$ matrix whose jth diagonal element $\Omega_{j,j} = \partial Q_j / \partial P_j$. For any choice of the parameters $\theta' = (\zeta', \gamma', \delta', \pi', \alpha')$, the corresponding vector of cost rates can thus be determined in closed form:

$$
\bar{c} = P + \Omega(P, X, W \mid \theta)^{-1} Q(P, X, W \mid \theta). \tag{11}
$$

The cost-rate residuals ϵ can then be determined for each outlet as the difference of the outlet's total cost rate and the average cost rate of the chain to which it belongs, i.e.,

$$
c_{k(j)} = \frac{1}{|J_k|} \sum_{j' \in J_K} \bar{c}_{j'}, \quad \epsilon_j = \bar{c}_j - c_{k(j)}, \quad \forall j = 1, \ldots, J. \tag{12}
$$

One might be tempted to estimate the unknown parameters from (11) with the help of standard maximum-likelihood methods. However, such methods require a choice of the specific unconditional distributions for the cost-rate noise terms ϵ. Moreover, because of the endogeneity of the price vector P, these variables are correlated with the noise terms, so that all conditional distributions $[\epsilon_j \mid P_l\colon 1 \leq j,\ l \leq J]$ need to be prespecified as well. Incorrect guesses for these various distributions result in biased inferences; see Hall (2005). The GMM technique overcomes both difficulties. See Nevo (2000) and Hall (2005) for clear expositions. It employs a vector of so-called instrument variables Z_j, which are correlated with (some of) the explanatory variables $\{P, X, W, D\}$, but uncorrelated with the cost-rate noise terms ϵ; i.e., $E[Z_j \epsilon_j] = 0$ for all cost rates and all outlets.

Our instruments are based on the four instrumental variables $V_{1j}, V_{2j}, V_{3j}, V_{4j}$ defined in §4. To account for asymmetries in the way that different chains are affected by these instrumental variables, we interact these variables with the chain indicator vectors, I^k, $k = 1, \ldots, K$, to arrive at a total of 24 instruments: for all $j = 1, \ldots, J$, Z_j is a (24×1) vector defined by $Z_j \equiv \{Z_{l, k, j} = V_{lj} \cdot I_{kj}, l = 1, \ldots, 4, k = 1, \ldots, K\}$. Intuitively, these instruments affect demand by altering the strength of competition and the size of the potential market. Moreover, they appear to be uncorrelated with the cost-rate differential an outlet is experiencing vis-à-vis the chain norm. Note that all of Cook County is urban or suburban with a total land area of only 946 square miles. It therefore faces a labor market with fairly uniform labor rates and skills. This implies that it is very unlikely that any cost efficiencies or inefficiencies of any outlet compared to the chain norm can be attributed to the population or worker density in its area. Had the study been conducted on a nationwide level, the assumption of independence would be more questionable. Other than labor cost, the remaining variable costs in this industry are associated with food, energy, and other process inputs that are tightly prescribed by the chains. Outlets may face different real estate costs depending upon whether

they are in downtown Chicago as opposed to suburban areas: however, these differences affect the outlets' fixed costs rather than their variable costs. In other words, it is reasonable to assume that the above-stated orthogonality conditions apply. In view of the population moment conditions, we must have, for the proper parameter vector θ, the sample average, $G(\theta) = (1/J) \sum_{j=1}^{J} Z_j \epsilon_j(\theta)$, of the vectors of random variables $\{Z_j \epsilon_j, j = 1, \ldots, J\}$ as close to zero as possible. The GMM estimator computes a parameter vector $\hat{\theta}$ that minimizes a quadratic function of this sample average; more specifically, for a given weighting matrix A,

$$\hat{\theta} = \arg \min_{\theta} G(\theta)' A G(\theta). \tag{13}$$

The optimal weighting matrix for the GMM estimator has been shown to be the inverse of the asymptotic variance-covariance matrix of the moment conditions. However, because this matrix is not available a priori, we follow the commonly used two-step estimation procedure: In the first step, we use the GMM with weighting matrix $A_1 = I$ to get a consistent initial estimator $\hat{\theta}_1$ from (13). We then use $\hat{\theta}_1$ to estimate the asymptotic variance-covariance matrix of the moment conditions, $(E[G(\hat{\theta}_1) G(\hat{\theta}_1)'])$, and solve the optimization problem (13) a second time with $A_2 = (E[G(\hat{\theta}_1) G(\hat{\theta}_1)'])^{-1}$ as the weighting matrix (see Hall 2005).

There are well-documented technical difficulties associated with optimization problem (13). Its objective function has many local optima. In addition, there are large regions where this function is close to flat, creating formidable difficulties for standard gradient methods. As a consequence, we designed a specific optimization method, described in Appendix B, and ran this algorithm with 20 different starting points.

Although there are asymptotically accurate approximations for the variances of the parameter estimates (see e.g., Hall 2005), these are often known to perform badly (see Brown and Newey 2002 and the 1996 special issue of the *Journal of Business and Economic Statistics* quoted therein). Therefore, to validate the statistical significance of these estimates, we have constructed confidence intervals using a bootstrapping procedure. This procedure is advocated when no sample data are available beyond those used to obtain the estimate (see, e.g., Brown and Newey 2002). The idea is to use subsets of the sample and calculate the value of the estimators in each subset in order to estimate the variance. To that end, we selected 80 random subsets of the tracts and ran the second stage of the above algorithm on each subset, for all 20 starting points, resulting in a total of 80 parameter vector estimates. Each subset has 134 tracts (10% of total number of tracts). Eighty subsets were used because

randomly selecting 80 out of 120 such subsets consistently yielded similar confidence intervals. We used the empirical distribution to construct the confidence intervals for each parameter.

We undertook two additional robustness tests for our estimates. To attempt to improve the efficiency of our estimates, we supplemented the 24 micromoments, introduced in §5, with additional so-called macromoments. Imbens and Lancaster (1994) suggest supplementing micromoments with macromoments to increase the efficiency of the estimates. This approach has been used in industrial organization studies by Petrin (2002) and Davis (2006). See Appendix C for a specification of the macromoments. In our second robustness check we ran the estimation using only subsets of the six chains included in our base model. We ran the estimation once assuming that only the three largest chains (McDonald's, Wendy's, and Burger King) have a presence in the market and a second time excluding the largest chain (McDonald's) from the market.

Finally, if a nonaffine capacity cost function is desired (see §3.2), a vector term $k(P, X, W \mid \theta)$ is to be added to the marginal profit expression $(P - \bar{c})$ in (10). It is easily verified that all steps in the estimation procedure continue to apply, with modest algebraic modifications. Whether the addition of nonlinear terms to the capacity cost function results in an upward or downward shift of the waiting time sensitivity depends on the coefficients in this four-parameter family of functions, themselves dependent on specific characteristics of the queueing system, e.g., the coefficients of variation of service and arrival time distributions (see Allon and Federgruen 2008).

6. Results

In this section, we report the results of the estimation process and robustness checks. We use either the chain's number of county outlets (OUT), the chain's national advertising spend (ADV), or neither of these, as an additional chain attribute X in the utility function (1). (The sensitivity to the ADV attribute was found to be statistically insignificant when this attribute was added as the single X-variable. We have therefore omitted the specification with both OUT and ADV as additional chain attributes.) We focus on the key parameters of interest, emphasizing those that are statistically significant. Table 3 reports the estimated value of each of the main demand coefficients: price, waiting time, and distance sensitivity (γ, δ, α), as well as the sensitivity to the additional chain attributes, if applicable. We report the estimates obtained by averaging all 20 two-stage optimization solutions as well as their 95% confidence intervals. The global optimum among all 20 two-stage solutions is consistently close

to the averages; for example, for the preferred specification with OUT as an additional chain attribute, the gap is never larger than 50%.

We first note that the estimates of the main parameters of interest, i.e., the sensitivity parameters for price, waiting time, and distance (α, γ, δ), are remarkably consistent across all three model specifications. This applies both to the point estimates and their 95% confidence intervals. (The only exception is the estimate of the distance sensitivitiy in the model without the additional chain attribute X, which is approximately five times smaller than the estimate in the other two specifications.) The best fit is obtained for the specification with OUT. Its coefficient is significantly positive at the 99% level, as opposed to that of the ADV variable, which is statistically insignificant. In view of the consistency of the estimates across the three model specifications and the relative superiority of the OUT specification, we focus on this specification in the remaining base model discussion as well as the robustness tests.

To ensure that the observed price vector is a Nash equilibrium under the estimated parameter values, we have verified that condition (C) is satisfied. Indeed, as proven under this condition in Theorem 3.1, we observed that $p_j < \bar{c}_j + 2/\gamma$ for all $j = 1, \ldots, J$. In fact, we found that $p_j < \tilde{p}_j = \bar{c}_j + 1/\gamma$, so that the observed price vector is in fact the *unique* equilibrium as long as $\tilde{p}$ is chosen as the maximum price vector; see the discussion at the end of §3 and Allon et al. (2009, Theorem 4.4).

Table 3 shows that all of the parameters $(\alpha, \delta, \gamma, \zeta)$ are significant at a 99% confidence level. Our estimates indicate that consumers attribute a very high cost to the time they spend waiting. Both the price and waiting time parameters have a significant impact on the consumer's decision. These results confirm our initial conjecture, as well as the belief expressed by industry experts that in the fast-food drive-thru industry customers trade off price and waiting time. In particular, to overcome an additional *second* of waiting time, an outlet will need to compensate an average customer by as much as $0.05 (=0.0237/0.492) in a meal whose typical price ranges from $2.25 to $6. This corresponds with an hourly cost rate of approximately 10 times the (pretax) average wage of $18/hour and nearly 30 times the (pretax) minimum wage in Illinois in 2005 ($6.50/hour). Even when comparing (opposite) extreme values of the 95% confidence intervals, the average consumer assigns a cost to waiting that corresponds with a rate of at least $0.04 per second. Because price differences in this industry, as in many others, are rather modest, this valuation implies that in the drive-thru market waiting time plays a more significant role than pricing in explaining sales volume. Moreover, these results seem to justify the continuing trend of chains making substantial investments to improve their waiting time.

It is also interesting to compare waiting time and distance sensitivity. After all, the disutility associated with the distance factor arises mainly from the associated time loss. Assuming, for example, an average velocity of 30 miles/hour, the estimate of δ implies that every additional second spent driving to the outlet reduces the utility measure by $0.824 \times 30/3{,}600 = 0.69E^{-2}$. Thus, the disutility of time spent waiting in the drive-thru line is at least three times that associated with the traveling time. This finding is consistent with the literature (see, e.g., Kahneman and Tversky 1984, Larson 1987) that reveals that individuals value time very differently, depending on the context and the degree to which time spent is pleasurable or not: most people mind time spent driving far less than time waiting idly; some even enjoy the ride.

There are some limitations to be noted with the estimation of the contribution of travel time to the overall utility value. First, we do not have an exact measure of the road distance between the consumer's residence and the outlets. Even if we did, this distance is not the best-possible measure for the additional effort and time she needs to expend to travel to the outlet. After all, many consumers stop in a drive-thru on the way from one point to the other, so that the disutility associated with travel time is not perfectly measured by the distance between the consumer's residence or work place and the outlet. Finally, it is not clear that a consumer's disutility from travel varies linearly with the distance driven.

When omitting waiting time as an explanatory variable in the consumer choice model, the resulting

Table 3   Estimates of Consumer Sensitivity to Key Attributes Under Three Model Specifications

Model specification	Price sensitivity ($)	Wait time sensitivity (sec.)	Distance sensitivity (miles)	Chain proxy sensitivity
No brand proxy	4.86E−01**	2.64E−02**	1.73E−01**	NA
	(4.50E−01, 4.97E−01)	(2.65E−02, 2.93E−02)	(1.27E−01, 2.22E−01)	NA
Number of outlets	4.92E−01**	2.37E−02**	8.24E−01**	1.39E−02**
	(4.05E−01, 5.00E−01)	(2.05E−02, 2.95E−02)	(8.11E−01, 1.26E+00)	(1.37E−02, 2.59E−02)
Advertising spend	4.92E−01**	2.34E−02**	9.15E−01**	−5.55E−04
	(4.12E−01, 5.02E−01)	(1.65E−02, 2.99E−02)	(8.69E−01, 1.29E+00)	(−1.24E−02, 3.01E−03)

**Indicates significance at the 99% confidence level. Significance level determined via a two-tail test.

price sensitivity estimate is 0.543, a 10.5% increase compared with the estimate we obtain with the full model. In fact, this estimate lies outside the confidence interval obtained using the full model. Indeed, when service-level attributes such as waiting time are disregarded, any reasonable estimation method can be expected to attribute a greater weight to price differentials to explain differences in sales volumes and market shares. Furthermore, estimating the elasticity of demand with respect to price using the model without waiting time results in overestimating this price elasticity by more than 10%. Not accounting for the waiting time as a strategic attribute also overstates the utility value of the outside good and disguises the importance of the chains' number of outlets in the county, the best identified indicator of the consumers' perception of chain quality. These various distortions contribute to suboptimal pricing decisions when ignoring waiting time as an explanatory variable in the consumer choice model.

It is also of interest to compare our results with those of Thomadsen (2005), who employed a similar model to estimate market share equations in the hamburger fast-food industry in Santa Clara, California. Thomadsen's consumer choice model disregards differences in service attributes as explanatory variables but includes the coownership structure (i.e., multioutlet owners) and brand dummies. Consistent with our findings regarding the impact of omitted service attributes, Thomadsen's estimate for the price sensitivity parameter is systematically larger than (in his case approximately double) the value we obtain. Along with considering all national chains with five or more outlets in the county, we represent the fast-food market as more competitive than Thomadsen does, in that we disregard the fact that a certain percentage of franchise owners own multiple outlets. (As mentioned, we lacked information about common ownership.) Ignoring the limited coownership phenomenon (see earlier discussion of Kalnins and Lafontaine 2004), results in underestimated equilibrium price sensitivity estimates for given observed price levels. Note that even if the price sensitivity parameters were double our estimate, the estimated cost of waiting time would be approximately $90/hour. Finally, Thomadsen (2005) reports only a 90% confidence interval on the price estimate, (0.14, 1.68), which has a margin of error nearly 20 times that of our 90% interval. Indeed, our confidence interval is entirely contained in his.

6.1. Robustness Testing

We have conducted various tests to confirm the robustness of our estimates beyond the consistency of the parameter estimates across the three model specifications, as well as the relatively narrow confidence

Table 4 Parameter Estimates Under Three Alternatives to the Base Model

	Base model	With MM	{McD, BK, WN}	Without McD
Price ($)	4.92E−01 (4.05E−01, 5.00E−01)	5.03E−01	5.08E−01	5.14E−01
Waiting time (sec.)	2.37E−02 (2.05E−02, 2.95E−02)	2.01E−02	2.05E−02	2.13E−02
Distance (miles)	8.24E−01 (8.11E−01, 1.26E+00)	7.70E−01	8.53E−01	7.16E−01
Brand proxy (no. of outlets)	1.39E−02 (1.37E−02, 2.59E−02)	9.07E−03	1.34E−02	2.47E−02

intervals. These additional robustness tests consist of (i) adding the macro moments discussed in §4 and listed in Appendix A to the GMM estimation procedure; (ii) repeating the estimation procedure under the assumption that only the three largest chains—McDonald's, Burger King, and Wendy's—have a presence in the county; and (iii) repeating the estimation under the assumption that the largest chain, i.e., McDonald's, is absent in the county, thereby reducing the number of outlets by 44%. All of the additional robustness tests employ the specification with the OUT attribute. Table 4 reports the estimates under the above three alternatives, while restating the estimates of the base model. (As before, the numbers within parentheses denote a 95% confidence interval; the estimates with macromoments are based on a single stage of estimation.) Once again, we have remarkable consistency in all of the parameter estimates among the different variants of the model/estimation procedure.

6.2. Counterfactuals

How much, then, is it worth to reduce the waiting time standards? We mentioned the industry maxim that a seven-second reduction in waiting times increases a chain's market share by 1%. We have therefore investigated the impact of a *single* chain reducing its waiting time standard by seven seconds, allowing all outlets to adjust their prices to the new price equilibrium. The results of this experiment can be seen in Table 5. In the first two rows (i.e., the rows with "Initial" in the title), we give the estimated daily demand and market share of each chain at the current waiting time standards and prices. The third row, titled "McD," shows the change in every chain's market share and demand volume when McDonald's reduces its waiting time by seven seconds. The following five rows contain the results of the same experiment for the remaining five chains. The percentage of the total market captured at the current waiting time standards closely matches the results in

Table 5   Change in Market Share Following Seven-Second Wait Standard Reduction at a Chain

	McD	BK	WN	Wh. Castle	DQ	S 'n S	Percentage of total market
Initial demand	5.58E+05	2.26E+05	1.33E+05	1.83E+04	1.95E+03	1.35E+03	14.22
Initial market share	59.47%	24.06%	14.16%	1.95%	0.21%	0.14%	
McD (Δ demand)	6.76E+04	−2.90E+03	−1.57E+03	−2.32E+02	−2.04E+01	−1.32E+01	15.17
(Δ market share)	3.02%	−1.80%	−1.05%	−0.15%	−0.02%	−0.01%	
BK	−2.93E+03	2.93E+04	−6.36E+02	−9.67E+01	−8.01E+00	−5.10E+00	14.61
	−1.88%	2.40%	−0.44%	−0.06%	−0.01%	0.00%	
WN	−1.58E+03	−6.38E+02	1.76E+04	−5.34E+01	−5.66E+00	−3.68E+00	14.45
	−1.12%	−0.45%	1.62%	−0.04%	0.00%	0.00%	
WC	−2.36E+02	−9.73E+01	−5.36E+01	2.48E+03	−6.47E−01	−2.43E−01	14.25
	−0.16%	−0.06%	−0.04%	0.26%	0.00%	0.00%	
DQ	−2.07E+01	−8.06E+00	−5.69E+00	−6.48E−01	2.64E+02	−7.12E−02	14.22
	−0.02%	−0.01%	0.00%	0.00%	0.03%	0.00%	
SS	−1.33E+01	−5.13E+00	−3.69E+00	−2.43E−01	−7.12E−02	1.84E+02	14.22
	−0.01%	0.00%	0.00%	0.00%	0.00%	0.02%	
All	6.23E+04	2.52E+04	1.51E+04	2.05E+03	2.25E+02	1.59E+02	15.81
	−0.01%	−0.01%	0.02%	0.00%	0.00%	0.00%	

Paeratakul et al. (2003), providing further validation of our estimates.

Our results confirm that the industry maxim is, on "average," correct. However, the absolute change in market share ranges from 3% at McDonald's (the market leader) to 0.04% at Dairy Queen, with Wendy's, the chain with the fastest service in 2007 and 2008, experiencing an increase by 1.33%. (The *percentage* increase in market share ranges between 4% at McDonald's and 20% at Dairy Queen.) Here, a chain's market share is defined as the chain's sales as a percentage of the total sales in the hamburger drive-thru industry. Even more importantly, an unmatched reduction of McDonald's waiting time standard by seven seconds results in an increase of its sales volume by approximately 15%. Note that the increase in demand comes primarily from attracting new customers to the market. The percentage of the potential fast-food market captured by all the chains grows by more than 1% when any of the three large players lower their waiting time. As further discussed in §7, any chain's unilateral waiting time reduction is likely to induce waiting time changes by the competing firms. Indeed, between 2005 and 2008, almost all chains gradually reduced their waiting time standards, McDonald's from 224 to 158 seconds and Wendy's from 173 to 131. Therefore, in the last row of Table 5 we report the impact of a simultaneous seven-second reduction of the average waiting time by *all* chains. This simultaneous service improvement results in the six chains capturing an additional 1.5% of the potential market. Relative market share changes are small, with Wendy's the prime beneficiary in relative terms, perhaps because, for it, the seven-second reduction is the largest relative service improvement among all six chains.

7.   Conclusions and Extensions

In this paper, we have proposed an approach to estimate how sales volumes for a service organization depend on all prices and waiting times of the various service providers in the region, along with other relevant attributes. We have applied this approach to the drive-thru fast-food industry in Cook County, Illinois. Here, consumers assign an implicit value to waiting in the drive-thru queue that amounts to many times the pretax U.S. wage, thus answering the first two of the four main research questions raised in §1. Most importantly, chains can improve their absolute and relative market shares very significantly by relatively modest reductions in waiting time, which explains why all chains make continuous efforts to shave off seconds from their consumer waiting time—reducing waiting time standards pays off handsomely in the fast-food industry. A seven-second reduction, the magnitude of Wendy's improvement from 2007 to 2008, implies an "average" increase of a chain's market share by approximately one percentage point, which confirms the above industry maxim and answers the third research question. However, for a large chain like McDonald's, it would result in an increase by more than 3% whereas the increase is 0.04% for a small chain like Dairy Queen, thus providing an answer to the fourth and final research question. The competitive dynamics are such that, to the extent feasible via incremental process and technological improvements, it is in all chains' interests to reduce their waiting times; this occurs to a

large extent because such service improvements result in more potential consumers selecting the fast-food option.

Several important extensions of our study and underlying model would be valuable. First, it is not clear whether the waiting time experience is best characterized by the average alone or by the addition of other measures such as the standard deviation and/or a percentile (say the 90th percentile) of the waiting time distribution. Even if the average waiting time is the best proxy, it is conceivable that the consumer's utility level diminishes in a nonlinear way with it. A similar nonlinear dependence on the distance variable may be explored as well. In addition, other service attributes such as the accuracy of the order-filling process and the clarity of the speaker and menu board could be included as explanatory variables in the random utility model (1).

Studying the impact of finer segmentations of the population, including past patronage of specific outlets, so as to estimate the impact of loyalty/inertia would be of interest. However, without sales volumes for the period of interest, this fine level of segmentation is not feasible. Nevertheless, in the modified specification of the utility functions, we have added explanatory variables that reflect brand penetration and awareness in Cook County.

Although, as explained in §4, we believe that a consumer choice model with a *single* price indicator per outlet is most appropriate for this industry, it could be worthwhile to test a far more detailed model that considers a menu of items to be purchased at every outlet and consumers choosing an outlet/menu item combination. It goes without saying that the data gathering and estimation challenges associated with such a detailed model are formidable, especially in the absence of sales data.

It would also be desirable to investigate how the chains in the industry select their waiting time standards and how the costs associated with waiting time reductions compare with the resulting revenue enhancements. To this end, it appears natural to view the price competition model in this paper as the second stage in a two-stage game preceded by a first-stage in which the chains as competing players select waiting time standards to maximize their profits. Specification and estimation of such a first-stage model meets with various challenges. First, Lafontaine and Shaw (1999) document that franchises typically pay a fixed periodic fee to the chain, along with a percentage of the revenues. However, it is unclear how these parameters are set as a function of the desired waiting time standard. It is also unclear how investments and operational costs depend on this strategic choice. Moreover, because chains select national standards, a two-stage game is needed with all U.S. outlets participating in the second-stage price competition game. It is also unclear whether all chains face the same potential lower limits for the waiting time standard. A final challenge is to verify whether the first-stage competition model has a (unique) pure-strategy equilibrium.

More broadly, the modeling approach and estimation technique of our study could be applied in other service industries in which consumers make purchasing decisions based on a steady-state service measure, as opposed to the one prevailing at the time they consider entering the service system. In most other service industries, one may expect that the service-level measures vary by individual service provider. This greatly simplifies the identification challenges in the specification of the consumer choice model, but increases the data collection effort because one needs direct observations for every provider in the chosen market. In addition, in our study we are able to estimate the model without knowledge of sales volumes or marginal cost rates because of the absence of unobservable firm-specific (i.e., outlet-specific) attributes that can be argued to have a consistent impact on consumers' purchasing decisions. In industries where such an argument is not valid, the estimation approach can only be applied if either sales volume or variable cost-rate data can be assessed.

Acknowledgments

Appendix A. Two Sample *t*-Tests
In this appendix, we report on the two-sample *t*-tests (assuming unequal variances) we conducted on all national waiting time observations for each of the six hamburger chains to verify whether the waiting time distributions vary by chain. The critical values for each test, with an alpha of 0.05, consistently rounded to 1.96. The *t* statistic is reported in the right-hand section of Table A.1.

Appendix B. The Optimization Routine
To mitigate the difficulties associated with the optimization problem (13), we restrict the feasible region for the parameter vector θ by imposing several reasonable constraints. For all chains $k = 1, \ldots, K$, let

$$con(\theta)_{b,m} = \sum_{j=1}^{J} S_{j,b,m}(P, W, X \mid \theta)/h(b,m), \quad b = 1, \ldots, B;$$
$$m = 1, \ldots, M$$

= the fraction of the population in tract b and socioeconomic group m that purchases a fast-food meal;

$\overline{con}(\underline{con})$ = an upper (lower) bound for the fraction of the population in any geographical area and any socioeconomic group to purchase a fast-food meal;

Table A.1   Two-Sample *t*-Test on National Chainwide Wait Time Observations

Chain	No. of obs.	Mean wait (sec)	Std. dev.	McDonald's	Burger King	Wendy's	White Castle	Dairy Queen	Steak 'n shake
McDonald's	598	224.27	151.38	—	4.09	6.66	−3.98	−0.70	−3.86
Burger King	600	192.28	116.69	−4.09	—	2.89	−7.25	−5.15	−7.70
Wendy's	596	173.34	109.81	−6.66	−2.89	—	−9.13	−7.92	−9.93
White Castle	334	269.45	173.83	3.98	7.25	9.13	—	3.57	0.55
Dairy Queen	528	230.10	128.18	0.70	5.15	7.92	−3.57	—	−3.40
Steak 'n Shake	328	262.69	141.22	3.86	7.70	9.93	−0.55	3.40	—

$J_k = \{j: k(j) = k\}$ denotes the set of outlets belonging to chain k;

$\hat{c}_k(\theta) = $ best estimate of chain k's standard cost rate where

$$c_k = (1/J_k) \sum_{j \in J_k} [P_j + \Omega(P, X, W \mid \theta)^{-1} Q(P, X, W \mid \theta)_j];$$

$$\underline{c}_k = 0;$$
$$\bar{c}_k = \min_{j \in J_k} P_j: k = 1, \ldots, K.$$

We impose the constraints:

$$\underline{con} \leq con(\theta)_{b,m} \leq \overline{con}, \quad \text{for all } b = 1, \ldots, B \text{ and}$$
$$m = 1, \ldots, M, \quad \text{(B1)}$$

$$\underline{c}_k \leq \hat{c}_k(\theta) \leq \bar{c}_k, \quad \text{for all } k = 1, \ldots, K.^3 \quad \text{(B2)}$$

Thus, instead of the unconstrained problem (13), we solve the constrained optimization problem: $(P) \min_\theta \{(13)$ s.t. (B1) and (B2)$\}$. To solve the constrained optimization problem, we replaced the soft constraints (B1) and (B2) by penalty functions that penalize any violations of these constraints. The penalty functions are multiplied with a common multiplier Λ, which within the course of our iterative algorithm is reduced sequentially to zero. More specifically, we have used the following perturbed objective:

$$G'(\theta) A G(\theta) + \Lambda \sum_b \sum_m \left\{ \log[\overline{con} - con(\theta)_{b,m}] \right.$$
$$+ \log[con(\theta)_{b,m} - \underline{con}]$$
$$\left. + \sum_k \log[\bar{c}_k - \hat{c}_k(\theta)] + \log[\hat{c}_k(\theta) - \underline{c}_k] \right\}. \quad \text{(B3)}$$

We have developed a special algorithm to solve (P) via the modified objective (B3). We begin with a large value for Λ, the weight of the penalty functions, roughly two orders of magnitude larger than the objective function value at the starting point. This is an application of the general barrier method approach for constrained nonlinear optimization. The algorithm invokes a quasi-Newton search method. During this search, we restrict movement in the

direction of the barriers imposed by the penalty functions so that any point within the interior of the feasible region can be reached, but points along the barrier are not approached very quickly, thus preventing the algorithm from "trapping" itself in unfavorable points. When a stopping condition is reached, the penalty weight Λ is halved and the modified quasi-Newton search rerun. In the first iteration, when the penalty Λ is large, this generally results in the algorithm moving to a point that is quite far from the barriers. The algorithm iterates until the penalty weight is small enough to render the penalty terms insignificant compared to the regular objective function (13). Because by the termination of the algorithm, the multiplier is reduced to an insignificant number, the algorithm optimizes the *true* objective function (13) over the feasible region described by the constraints (B1) and (B2).

To arrive at the reported estimates, we used a process in which, in the first stage, we took 20 starting points and ran the above algorithm with two different initial values of the penalty parameter Λ—one is two orders of magnitude larger than the other—resulting in two estimates per starting point. For each of the 20 starting points, we chose the estimate (of the two) that resulted in the lower objective function (excluding the penalty function) and generated a weighting matrix for this estimate from the covariance matrix. In the second stage we ran our algorithm starting with this estimate and weighting matrix, again from both Λ values generating 40 final estimates.

Appendix C. The Macromoments

We have added macromoments that are based on three demographic features: age, race, and gender. We use the study by Paeratakul et al. (2003), which reports the proportion of people in various demographic groups that consume fast food over a two-day period. As suggested in Thomadsen (2005), the macromoments are constructed based on the idea that the consumption ratio of related demographic groups in Cook County should be close to the national consumption ratios. For example, the local ratio of men to women consuming a fast-food meal should match the national ratio; i.e., the percentage of women consuming fast food in Cook County may differ from the national average, but the fraction of men consuming should differ from the national average proportionally to women. The following 12 macromoments were added to the micromoments based on comparisons between age brackets, one between genders, and one between races.

---

[3] We have chosen the constraints 0.05 and 0.70 as a wide range around the national average consumption percentages reported in Paeratakul et al. (2003). In this study, which surveyed a national sample of the population across various age, gender, and race combinations to see if they had eaten fast food over the previous two-day period, the group with the highest (lowest) percentage of fast-food consumption reported at 52% (20%), well within our bounds.

$$G_{0-9,10-16}(\theta) = \frac{1}{J} \sum_{j=1}^{J} \left[R_{0-9} \frac{Q_{j,10-19}(\theta)}{Pop_{10-19}} - R_{10-19} \frac{Q_{j,0-9}(\theta)}{Pop_{0-9}} \right],$$
$$\text{(C1)}$$

$$\cdots$$

$$G_{40-59,\,60+}(\theta) = \frac{1}{J}\sum_{j=1}^{J}\left[R_{40-59}\frac{Q_{j,\,60+}(\theta)}{Pop_{60+}} - R_{60+}\frac{Q_{j,\,40-59}(\theta)}{Pop_{40-59}}\right],$$

(C2)

$$G_J(\theta) = \frac{1}{J}\sum_{j=1}^{J}\left[R_{\text{Male}}\frac{Q_{j,\,\text{Female}}(\theta)}{Pop_{\text{Female}}} - R_{\text{Female}}\frac{Q_{j,\,\text{Male}}(\theta)}{Pop_{\text{Male}}}\right], \quad \text{(C3)}$$

$$G_J(\theta) = \frac{1}{J}\sum_{j=1}^{J}\left[R_{\text{Black}}\frac{Q_{j,\,\text{White}}(\theta)}{Pop_{\text{White}}} - R_{\text{White}}\frac{Q_{j,\,\text{Black}}(\theta)}{Pop_{\text{Black}}}\right], \quad \text{(C4)}$$

where R_{0-9} denotes the national fraction of fast-food consumers who belong to the 0–9 age bracket as estimated by the Paeratakul et al. (2003) study, Pop_{0-9} denotes the Cook County population in this age bracket, and $Q_{j,\,0-9}(\theta)$ denotes the demand of consumers age 0–9 at outlet j. Similar definitions pertain to the other $R\_$, $Pop\_$, and $Q\_$ numbers.

References

Allon, G., A. Federgruen. 2007. Competition in service industries. *Oper. Res.* **55**(1) 37–55.

Allon, G., A. Federgruen. 2008. Service competition with general queueing facilities. *Oper. Res.* **56**(4) 827–849.

Allon, G., I. Gurvich. 2010. Pricing and dimensioning competing large-scale service providers. *Manufacturing Service Oper. Management* **12**(3) 449–469.

Allon, G., A. Federgruen, M. Pierson. 2009. Price competition under multinomial logit demand functions with random coefficients. Working paper, Kellogg School of Management, Northwestern University, Evanston, IL.

Berry, S. T. 1994. Estimating discrete-choice models of product differentiation. *RAND J. Econom.* **25**(2) 242–262.

Berry, S., J. Levinsohn, A. Pakes. 1995. Automobile prices in market equilibrium. *Econometrica* **63**(4) 841–890.

Bresnahan, T. F. 1987. Competition and collusion in the American automobile industry: The 1955 price war. *J. Indust. Econom.* **35**(4) 457–482.

Brown, B. W., W. K. Newey. 2002. Generalized method of moments, efficient bootstrapping, and improved inference. *J. Bus. Econom. Statist.* **20**(4) 507–517.

Cachon, P. C., P. T. Harker. 2002. Competition and outsourcing with scale economies. *Management Sci.* **48**(10) 1314–1333.

Chen, H., Y.-W. Wan. 2003. Price competition of make-to order firms. *IIE Trans.* **35**(9) 817–832.

Davis, M. M., T. E. Vollmann. 1990. A framework for relating waiting time and customer satisfaction in a service operation. *J. Services Marketing* **4**(1) 61–69.

Davis, P. 2006. Spatial competition in retail markets: Movie theaters. *RAND J. Econom.* **37**(4) 964–982.

Deacon, R. T., J. Sonstelie. 1985. Rationing by waiting and the value of time: Results from a natural experiment. *J. Political Econom.* **93**(4) 627–647.

DeHoratius, N., M. Fisher, S. Netessine. 2010. McDonald's corporation: Launching McCafe. *Consortium for Operational Excellence in Retailing Conf. The Wharton School, University of Pennsylvania, Philadelphia.*

De Vany, A. S., W. L. Gramm, T. R. Saving, C. W. Smithson. 1983. Production in a service industry using customer inputs: A stochastic model. *Rev. Econom. Statist.* **65**(1) 149–153.

Feenstra, R., J. Levinsohn. 1995. Estimating markups and market conduct with multidimensional product attributes. *Rev. Econom. Stud.* **62**(1) 19–52.

Fisher, M. 2007. Strengthening the empirical base of operations management. *Manufacturing Service Oper. Management* **9**(4) 368–382.

Gans. N., G. Koole, A. Mandelbaum. 2003. Telephone call centers: Tutorial, review and research prospects. *Manufacturing Service Oper. Management* **5**(2) 79–141.

Garber, A. 2005. Quick-service leaders eye life in the fast lane: Chains try to increase order accuracy and shave valuable seconds off drive-thru times. *Nation's Restaurant News* (December 12).

Hall, A. R. 2005. *Generalized Method of Moments*. Oxford University Press, New York.

Hasija, S., E. J. Pinker, R. A. Shumsky. 2008. Call center outsourcing contracts under information asymmetry. *Management Sci.* **54**(4) 793–807.

Hassin, R., M. Haviv. 2003. *To Queue or Not to Queue: Equilibrium Behavior in Queueing Systems.* Kluwer Academic Publishers, Boston.

Hess, S., M. Bierlaire, J. W. Polak. 2005. Estimation of value of travel-time savings using mixed logit models. *Transportation Res. Part A.* **39**(2–3) 221–236.

Hughlett, M. 2008. Drive-throughs done right ring up returns. *Chicago Tribune* (November 28), http://articles.chicagotribune .com/2008-11-28/news/0811270365_1_drive-through-restaurant -technologies-competitive-advantage.

Imbens, G. W., T. Lancaster. 1994. Combining micro and macro data in microeconometric models. *Rev. Econom. Stud.* **61**(4) 655–680.

Jargon, J. 2006. McD's service stalls at drive-thru. *Crain's Chicago Bus.* (January 2).

Kahneman, D., A. Tversky. 1984. Choices, values, and frames. *Amer. Psych.* **39**(4) 341–350.

Kalnins, A., F. Lafontaine. 2004. Multi-unit ownership in franchising: Evidence from the fast-food industry in Texas. *RAND J. Econom.* **35**(4) 747–761.

KC, D. S., C. Terwiesch. 2009. Impact of workload on service time and patient safety: An econometric analysis of hospital operations. *Management Sci.* **55**(9) 1486–1498.

KC, D. S., C. Terwiesch. 2011. An econometric analysis of patient flows in the cardiac intensive care unit. *Manufacturing Service Oper. Management*, ePub ahead of print September 2. doi:10.1287/msom.1110.0341.

Lafontaine, F., K. L. Shaw. 1999. The dynamics of franchise contracting: Evidence from panel data. *J. Political Econom.* **107**(5) 1041–1080.

Larson, R. C. 1987. Perspectives on queues: Social justice and the psychology of queueing. *Oper. Res.* **35**(6) 895–905.

Levhari, D., I. Luski. 1978. Duopoly pricing and waiting lines. *Eur. Econom. Rev.* **11**(1) 17–35.

Luski, I. 1976. On partial equilibrium in a queuing system with two servers. *Rev. Econom. Stud.* **43**(3) 519–525.

Magid Associates. 2010. Drive-thru satisfaction: Presentation to *QSR* magazine. Report, Magid Associates, Minneapolis. http://www.qsrmagazine.com/reports/drive-thru-experience.

Mueller, C. D. 1985. Waiting for physicians' services: Model and evidence. *J. Bus.* **58**(2) 173–190.

Musalem, A., M. Olivares, E. T. Bradlow, C. Terwiesch, D. Corsten. 2010. Structural estimation of the effect of out-of-stocks. *Management Sci.* **56**(7) 1180–1197.

Nevo, A. 2000. A practitioners guide to estimation of random coefficients logit models of demand. *Econom. Management Strategy* **9**(9) 513–548.

Nuckolls, M. W. 2007. Building a better drive-thru: Consumer survey—The results of the 2007 *QSR* consumer drive-thru preference study. *QSR* (October), http://www.gokis.net/self-service/archives/2007_QSR_Drive_Thru_Preferences.pdf.

Olivares, M., G. P. Cachon. 2009. Competing retailers and inventory: An empirical investigation of General Motors' dealerships in isolated U.S. markets. *Managment Sci.* **55**(9) 1586–1604.

Olivares, M., C. Terwiesch, L. Cassorla. 2008. Structural estimation of the newsvendor model: An application to reserving operating room time. *Management Sci.* **54**(1) 41–55.

Paeratakul, S., D. P. Ferdinand, C. M. Champagne, D. H. Ryan, G. A. Bray. 2003. Fast-food consumption among U.S. adults and children: Dietary and nutrient intake profile. *J. Amer. Dietetic Assoc.* **103**(10) 1332–1338.

Petrin, A. 2002. Quantifying the benefits of new products: The case of the minivan. *J. Political Econom.* **110**(4) 705–729.

Png, I. P. L., D. Reitman. 1994. Service time competition. *RAND J. Econom.* **25**(4) 619–634.

So, K. 2000. Price and time competition for service delivery. *Manufacturing Service Oper. Management* **2**(4) 392–409.

Thomadsen, R. 2005. The effect of ownership structure on prices in geographically differentiated industries. *RAND J. Econom.* **36**(4) 908–929.