

# Does it pay to reduce your customers' wait? An empirical industrial organization study of the fast-food drive-thru industry based on structural estimation methods.

Gad Allon

Kellogg School of Management, 2001 Sheridan Road Evanston , IL 60208 , g-allon@kellogg.northwestern.edu

Awi Federgruen

Columbia, Graduate School of Business, af7@columbia.edu

Margaret Pierson

Columbia, Graduate School of Business, mpierson05@gsb.columbia.edu

October 29, 2009

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, what we believe to be the *first*, empirical study of an important industry to test whether and to what extent waiting time performance measures impact 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

The objective of this paper is to conduct an empirical study of an important industry to test whether and to what extent waiting time performance measures at competing firms impact these firms' market shares and price decisions. Most executives realize that time is money for the consumer, but it is unclear how much money and how the exchange rate differs in different industries and how it varies with other factors such as location, brand etcetera. For example, for the drive-thru fast-food industry, we show that customers attribute an implicit value to their wait time, which is anywhere between \$40/hr and \$180/hr. Similarly, transportation research studies such as that of Hess et al. (2005) estimate the implicit cost associated with travel time to be in excess of \$100/hr. Therefore, to assume that customers attribute a uniform cost rate to their waiting time and that this uniform rate equals, say, the per capita income in the market may result

in questionable policy analyses; see e.g. Mueller (1985). We also characterize how the price equilibrium responds to changes in the waiting standards. Additionally, we wish to understand whether the trend in various industries to continuously improve waiting times and service levels can be explained on (game-)theoretical grounds. To answer these questions, a competition model, derived from underlying consumer choice models, has been estimated using detailed empirical data. While our empirical study is focused on the drive-thru fast food industry, we introduce a methodological framework which, *mutatis mutandis*, can be employed to establish the impact of service measures on market shares in other industries.

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. In their popular textbook "Competing Against Time", Boston Consulting Group partners Stalk and Hout (1990) documented how time based competition has been reshaping global markets. Often, specific waiting time standards or guarantees are advertised. For example, Ameritrade has increased its market share in the on line discount brokerage market by "guaranteeing" that trades take no more than 10 seconds to be executed; the guarantee is backed up with a complete waiver of commissions in case the time limit is violated. This has led most major on line 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. See Allon and Federgruen (2007) for a longer list of examples. In other industries, average waiting times are monitored by independent organizations; examples include the airline industry where independent government agencies (e.g., the Aviation Consumer Protection Division of the Department of Transportation) as well as Internet travel services (e.g., Expedia) report, on a flight by flight basis, the average delay and percentage of flights arriving within 15 minutes of schedule. Another example is the fast food industry, the prime focus of this paper, where industry trade organizations such as Quick Service Restaurants, publicize yearly surveys of the average waiting time experienced at the various fast food chains. While the vast majority of fast food outlets are owned by independent franchisees, and while these franchisees select their own prices, a common waiting time standard is chosen by each of the chains. Chains prescribe a uniform operational process to their franchisees along with specific recipes for their standard menu items. These processes include the order taking process, standard customer greetings,

the number and the maximum amount of time burgers can be on the grill, the relationship between lanes and demand volume et cetera, all of which determine the chain's waiting time standard; see Garber (2005) and Jargon (2006). As a consequence, chains achieve remarkably uniform average waiting times at their franchises<sup>1</sup>. Moreover, chains invest heavily to shave seconds off the average waiting times, clearly believing that their market shares are very sensitive to the relative waiting times experienced.

Beginning with the seminal paper by Naor (1969), stochastic models for service systems have been analyzed in which the expected demand volume depends on either the steady state waiting time or the specific waiting time to be experienced by a customer seeking service. Additional early examples in the economics literature include Luski(1976), Levhari and Luski (1978), De Vany and Saving (1983), and De Vany et al. (1983). In the operations management literature, it has been conventional to represent a service system's demand process as an exogenous stochastic process, the characteristics of which do not depend on the waiting time experienced or any other controllable attribute such as the price charged. However, starting with Lee and Cohen(1985), Mendelson (1985) and Mendelson and Whang(1990), (queueing) models have started to acknowledge that potential customers, even in a monopoly setting, are able to choose whether to buy the service or not. In particular, the demand volume for a given service is to be represented as a *decreasing* function of, for example, the average waiting time or other characteristics of the steady state waiting time distribution. This *downward* sloping demand curve is to be distinguished from the *upward* sloping supply curve manifest in all queueing models showing that, say, the average waiting time is an increasing function of the customer (demand) volume. It is therefore problematic to estimate the relationship between customer volume and average waiting time from a simple regression, as has been undertaken, for example, for dentist practices, in De Vany et al. (1983). In the presence of multiple service providers, consumers face the additional choice of whom to patronize and select by trading off relative waiting times, along with prices and other attributes.

More recently, a series of theoretical models, for example Lederer and Li (1997), Cachon and Harker (2002) and Allon and Federgruen (2007) have been proposed to analyze a general market for an industry of

<sup>1</sup> As an example in Jargon (2006), Jim Hyatt, Burger King's vice president and chief global operations officer, states that the average waiting time at its 6,900 domestic franchises cover a very narrow range from 165 to 170 seconds, in 2006, while this average waiting time was reduced by 22 seconds, compared to the previous year.

competing service providers. Firms differentiate themselves by their price levels and the waiting time their customers experience, as well as different attributes not determined directly through competition such as the convenience of the pick-up process, or the accuracy with which the service process is delivered. Customers select a specific firm by trading off these three categories of service attributes.

To date, almost all contributions to the literature on service competition have been theoretical, with numerical investigations confined to small hypothetical examples<sup>2</sup>. Indeed, we believe ours to be the first empirical study to complement this theoretical literature on service competition. There are several reasons, to our knowledge, why a voluminous array of theoretical models has not been complemented by any empirical studies. First, 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. Yet, while 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 which, along with the firms' other strategic choices, explain ultimate consumer choices and hence, realized market shares. Second, it is, typically very hard, if not impossible, to collect data on sales volumes by the competing firms. Such data are sometimes accessible for consumer products, such as automobiles, pharmaceuticals or consumer goods sold in supermarkets (In the automobile industry, such data are generally, although not universally, available at the dealer level, because cars need to be registered and registration data are publicly accessible; similarly, every prescription has a prescription number which is entered into a nationwide data base, and which identifies the prescribing physician. As to goods sold at supermarkets, marketing firms like IRI and Nielsen sell sales data collected from scanners). In contrast, in the service industry, it is even rarer that sales volumes can be gathered and firms are reluctant to provide the information, considering it of the highest strategic value. Indeed, sales

<sup>2</sup>To our knowledge, the only empirical studies in this area are Deacon and Sonstelie (1985) and Png and Reitman (1994). The former estimates the joint impact the price differential and difference in average waiting time has on the sales at two neighboring gas stations in Ventura, California. The price differential occurs because of temporary price controls affecting chain owned stations differently from franchised stations. Thus, unlike almost all other industry settings, the price differential is exogenously determined, simplifying the estimation process greatly, as explained below. Png and Reitman (1994) studies what impact average waiting times have on peak hour retail gasoline sales using all 1501 stations in 4 Massachusetts counties. However, that study is not based on direct waiting time observations, but on a proxy for waiting times postulated as a specific function of weakly sales, number of operating hours and number of pumps. Moreover, Png and Reitman (1994) do not estimate how the competing firms' price decisions depend on each other and their waiting times, so that the estimates can not be used to predict the impact changes in waiting time standards would have on the firms' market shares and prices.

volumes were unavailable in our context as well. Instead, we were able to infer them by estimating the parameters in the system of equations characterizing the unique equilibrium in the above mentioned competition model. In other words, instead of computing the equilibrium from a set of estimated demand functions and cost structures, the demand functions 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.

To conduct our empirical tests we have selected the drive-thru fast-food industry. The fast-food industry realized over a hundred billion \$ in sales in 2007 with hamburger sales representing 73% of the market. McDonald's alone reported in excess of \$28 billion in revenues, representing 46% of the quick service burger market. The drive-thru sector accounts for about 70% of the industry's sales, a 10% increase from 6 years ago Hughlett (Nov 28, 2008). As reported by QSR (Quick Service Restaurant) Magazine, the aforementioned industry's trade publication, firms invest heavily to improve customer waiting and service times and the accuracy with which orders are filled. As far as the former is concerned, seconds are perceived to matter significantly. In the most recent (2008) survey of QSR of the top 25 chains, it was found that the best performing chain in terms of average waiting times (Wendy's) was twice as fast as the least performing chain, 13% better than the second performing chain (Bojangles') and 17% better than McDonald's, the industry leader. Moreover, Wendy's improved their waiting time by over 3% from the QSR survey results 3 years earlier. Many fast food chains have installed timer systems. These let an operator know how many cars visited the drive-thru at various times of the day, and, in addition, the average time customers spend in the drive-thru and which point had the longest wait time. Consumer surveys by QSR focus on the maximum waiting time customers are willing to tolerate. Recently, in a Chicago Tribune article, R. Craig Coulter, chief scientist at HyperActive Technologies, a restaurant software firm, stated that "People decide whether to come to your restaurant based on how long the drive-thru line is" Hughlett (Nov 28, 2008). In the same article, the president of data management at Restaurant Technologies, states that there's an industry maxim that for every seven-second reduction in drive-thru service time, sales will increase 1% over time.

We report on a large scale empirical industrial organization study in which the demand equations for restaurants are estimated based on so-called structural estimation methods. As mentioned, the demand equations are inferred from the equations characterizing the unique equilibrium in the competition model which results from a detailed consumer choice model and a cost structure reflecting a broad category of queueing systems. More specifically, as in most service industries, it appears impossible to obtain outlet specific sales data because the firms keep this information as propriety data. Building on the framework of Feenstra and Levinsohn (1995) and Berry et al. (1995), we accommodate the absence of demand data with three assumptions: (1) The consumers attribute a utility level to each potential outlet, which depends stochastically on price, waiting time, the distance to the outlet and various other outlet specific characteristics; similarly, customers assign a utility level to the no-purchase option, which depends stochastically on the consumer's gender, race, age bracket and occupational status. (2) Firms encounter a cost structure which 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. (2) Firms adopt the (unique) Nash equilibrium in the price competition model which 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 firms' cost structure. Our estimation method is a Generalized Method of Moments (GMM), as opposed to more standard maximum likelihood estimators for systems of non-linear equations. This is done to avoid making specific distributional assumptions about the error terms, and to circumvent the fact that these error terms are correlated with the explanatory variables. The data we use consist of the prices, service levels, locations and outlet distributions of the top fast-food hamburger chains in Cook County, Illinois. To accurately represent the impact travel distances and demographic features may have on the consumer choices, we divide the county into more than 1300 so-called tracts, a geographic unit used by the U.S. Census Bureau, with an average area of only 1.2 square miles, and employ the demographic composition of the population in each tract.

We estimate the model parameters and, in particular, focus on the impact that pricing and waiting decisions have on the demand for the products at each location. We also focus on how various environmental factors (such as the restaurant location and the chain identity) impact on the equilibrium choices. We use the fitted models to conduct counter-factual experiments that demonstrate how the geographic, brand identity and demographic properties jointly determine the prices and service levels. Some of the principal insights obtained are that customers attribute a large cost to waiting, and that differences in waiting time standards play a more significant role in explaining differences in sales volumes than price differentials. When waiting time standards and prices are determined sequentially in a two stage competition game, we observe that the competitive dynamics drive the firms to reduce waiting time standards to their minimum levels, as dictated by organizational or technological considerations.

In summary, the main contribution of this paper is that, to our knowledge, it is one of the first to estimate how sales volumes for a service organization depend on all price and waiting times of all competing providers within a reasonable geographic distance, as well as other attributes (e.g, brand-specific characteristics). In particular, we use structural estimation methods to study how customers trade off prices and waiting times in an empirical market-based study, concluding for the fast food industry that consumer attribute a value to their wait time of at least \$40/hr. We confirm that a 7 second reduction results, on “average”, in a 1% market share increase. However, for a large chain like McDonald’s, the increase is by more than 3% and sales go up by 15%. Our model explains the continuing trend of all chains investing heavily to reduce their waiting time standards.

The remainder of the 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 them. Section 5 is devoted to a description of the Generalized Moment Method as applied to our model. Section 6 describes the estimation results and completes the paper with a discussion of the above mentioned counter-factual studies.

## **2. Literature Review**

The literature on competition in service industries dates back to the late 1970s. As mentioned in the introduction, 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 given 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 whether a price equilibrium exists in this model remained an open question, until, for the basic model with a uniform cost rate, it was recently resolved in the affirmative by Chen and Wan (2003). These authors show, however, that the Nash equilibrium may fail to be unique. More recent variants of the Levhari and Luski models include Li and Lee (1994), Armony and Haviv (2001) and Wang and Olsen (2008).

Loch (1991) considers a variant of the Luski model in which the service times of the two providers have a *general*, though still identical, distribution, i.e. in which each provider is modeled as an M/G/1 system. Assuming that the total demand rate for service is given by a general function of the *full price*, the author shows that a symmetric equilibrium pair of prices exists, irrespective of whether the two firms target prices *directly* (Bertrand competition), or *indirectly*, via demand rates (Cournot competition). Lederer and Li (1997) generalize Loch (1991) to allow for an arbitrary number of service providers and a finite number of customer classes, each with a given waiting cost rate.

In the above papers, firms compete in terms of their price (only), with fixed exogenously specified capacity levels (§7 in Chen and Wan relaxes this assumption; see below). Several other papers assume, alternatively, that prices are fixed while firms compete in terms of their capacity levels. Kalai et al. (1992) consider, again, a duopoly with Poisson arrivals and exponential service times. A fixed customer population joins, upon arrival, a single queue from which they are served on a FIFO basis by the first available server. (When a customer arrives to an empty queue, he is randomly assigned to one of the two providers). In this model asymmetric Nash equilibria of service rate pairs may arise. Gilbert and Weng (1998) and Cachon and Zhang (2007) extended the above model to allow for routing probabilities that depend on the providers' service rates according to more general (allocation) schemes.

De Vany and Saving (1983) are the first to address a richer type of competition in which firms compete with *several* rather than a *single* strategic instrument. This paper addresses a variant of the Levhari and Luski model with an arbitrary number of identical firms who *simultaneously* choose a *price* and *service*

*rate*. All customers share the same waiting cost rate, but the total demand volume in the industry is given by a general function of the lowest *full price*. The authors establish the existence of a symmetric equilibrium.

Cachon and Harker (2002) and So (2000) analyzed the first models in which customers consider additional criteria beyond the lowest full price when choosing a service provider. Both confined themselves, again, to M/M/1 service providers. Cachon and Harker (2002) considered the case of *two* firms where demand rates are given as either *linear* or MultiNomial Logit functions of the two full prices. So (2000) considers an arbitrary number of competing firms and a different class of so-called attraction models. In an attraction model, each firm's market share is proportional to an attraction value, specified as a function of several of the firm's attributes. The competition model derived and estimated in this paper is a generalization of the attraction model, in which the parameters in the attraction value functions are treated as random. In So (2000), the logarithm of the firms' attraction value is specified as a common linear combination of the logarithm of the price and the logarithm of the waiting time standard, plus a firm dependent constant. This specification continues to imply that the price and waiting time are aggregated into a single, albeit, different *full price* measure. Allon and Federgruen (2006,2007) appear to be the first models to treat the price and waiting time standard as completely independent firm attributes which different customers may trade off in different ways. Nevertheless, Allon and Federgruen (2007)'s paper confined itself to systems of demand rates that are linear in the prices and to M/M/1 service providers while Allon and Federgruen (2006) studies more general demand models such as attraction models and allow for more general queueing facilities. We refer to these two papers as well as Hassin and Haviv (2003) for a recent survey text on queueing models with competition.

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. (2007) for a recent survey of such agreements. However, virtually all planning models in this vast literature 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. When describing future challenges in this area, the authors single out determining "a better understanding of customer behavior" (§7.3) and the need to model and estimate "multiple levels of equilibria". Beyond the levels mentioned

there, we suggest the desirability of models incorporating the competition effects of service levels provided by the call centers of competing service providers. Another stream of papers, in particular Hall and Porteus (2000) and Gans (2002), models the competition between services providers selecting a distribution for the (non-congestion related) quality of service, based on specific consumer choice models. Gans et al. (2007) describes an empirical study to test these models, based on laboratory experiments, as opposed to econometric field studies, as in this paper. In the fast-food industry, Davis and Vollmann (1990) recorded 723 customer waiting times and asked these customers 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 (eg - shopping, visiting friends) were not significant. This empirical study complements the above quoted plethora of trade literature documenting the centrality of waiting times in this industry.

The above reviewed literature is based on the observation that firms compete along the service level dimension as well as anecdotal and the above summarized empirical evidence that customers value waiting time when making decisions regarding their preferred service provider. Our study complements this literature by estimating the parameters used and assumed by these models. As mentioned, the two empirical studies closest to ours 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 ours. The selected estimation method is based on a probit model, applicable in the case of two firms only; 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) describes the peak hour sales in a market of 1501 gas stations in 4 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 in 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 as well as 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, while 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 which 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. Similar to our paper, the transportation research literature has often specified the demand for alternative transportation modes as arising from a Mixed Multi Nomial Logit model with prices and travel times as explanatory variables for the utility measures. The maximum likelihood estimations typically employed are, however, challenged by the above mentioned endogeneity problems, see Hess et al. (2005) for a recent example.

The approach we use to estimate the impact of waiting times, prices, geographic dispersion, chain attributes, and socio-economic factors on demand follows the work by Bresnahan (1987), Berry (1994), Berry et al. (1995). 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. (Berry et al. (1995) applied this approach to study price competition in the US automobile industry.) The estimation method, based on the (GMM) allows for prices to be determined endogenously (as the equilibrium of an underlying competition model), rather than being selected exogenously. It has the additional advantage of avoiding distributional assumptions for the error terms in the equations to be estimated. 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 since in the fast food industry, sales data are not reported and treated as strategic and proprietary information. See also Dube et al. (2008).

More recent work by Davis (2006) and Thomadsen (2005a) incorporated geography in the BLP framework. More specifically, Thomadsen (2005a) studies the impact of ownership structure on prices in the fast food industry. Thomadsen (2005a) 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 measures to the set of outlet and chain-dependent explanatory variables employed in Thomadsen (2005a) so as to characterize and estimate how outlets and chains use the combination of price and waiting time standards in their competitive battles.

Our study is related to the recent empirical literature in operations management. To our knowledge, with the notable exception of Olivares et al. (2008), most of this literature focuses on consumer products rather than services. See Chen et al. (2005), Chen et al. (2007), Gaur et al. (2005), Olivares and Cachon (2007), Musalem et al. (2009) and the references therein for surveys of this literature.

### **3. The Model**

In this section we develop the competition model representing the competitive interdependencies and interactions between the outlets in our geographic region (i.e. Cook county). The model combines two sub-models

(a) a consumer choice model which determines how many of the residents of the region 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 its sales volume and service level, i.e. its waiting time standard.

Combining the two sub-models permits us to derive the outlets' profit functions. We then characterize the equilibrium behavior in the price competition model under given waiting time standards, as specified by six major chains operating in the selected geographical region. We show that the model has a unique equilibrium which is the unique solution of a non-linear system of equations. It is this system of equations which 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. We conclude the section with a description of a plausible

model for the joint determination of *all* outlets' prices and all chains' waiting time standards. The model embeds the aforementioned price competition model as the second game (with the outlets as independent "players") of a two-stage game, where the first stage game has the chains as competing players selecting the waiting time standards. As mentioned in the introduction, in the fast food industry, waiting time standards are selected and prescribed by the chains: however, price decisions are relegated to the independent outlets to avoid illegal forms of price fixing, if for no other reason. As franchising became popular in the sixties, the US courts began to limit the types of pricing restrictions chains can impose on their franchises. Even the specification of maximum retail prices has become illegal, by the Supreme Court ruling in *Albrecht vs. Herald* (1968). (Indeed, we have observed significant price differences among outlets of the same chain, see Table 3 in Section 4.) In general, one may envision other ways in which the two sets of strategic choices (price and waiting time standard) are selected: for example, settings where all choices are made simultaneously and those where prices are selected in advance of waiting time standards. See Allon and Federgruen (2007) for a systematic comparison of the three types of competition models, in the context of demand functions that are linear in the prices. Indeed, these authors show that the equilibria arising in these three models are, in general, different and, under mild conditions, more aggressive service levels emerge in equilibrium when set in advance of prices, as opposed to being chosen simultaneously with the prices. However, in the fast-food, only the waiting time first-price second competition model is applicable.

### **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, observable (to the modeler) attributes 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 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 which depends on his or her demographic attributes. The customer also assigns a utility value to each of the outlets in the region that depends on the attributes of the outlet and the chain it belongs to. 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:

$$U_{i,j} = X_j' \beta - D_{ij} \delta - P_j \gamma - W_j \xi + \eta_{ij} \quad (1)$$

where  $X_j$  is a vector of dummy variables indicating the chain identity of the outlet (as well as, possibly, other observed properties),  $D_{ij}$  is the distance between consumer  $i$  and outlet  $j$ ,  $P_j$  is the price of a (standard) meal at outlet  $j$ , and  $W_j$  is the waiting time standard associated with outlet  $j$ ;  $\beta, \delta, \xi$  and  $\gamma$  are parameters to be estimated, and  $\eta_{ij}$  is the unobserved (to the modeler) portion of the utility of individual  $i$  at outlet  $j$  arising from factors such as the cleanliness of the outlet or the safety of the location.

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 vector specifying the consumer's age, gender, race, and whether they are at work or at home when making the decision (occupational status);  $\beta_0$  and  $\pi$  represent another set of parameters to be estimated and  $\eta_{i0}$  denotes the unknown portion of the utility of individual  $i$  for the non-purchase option.

We distinguish among a limited number of age brackets. Therefore, there is a finite list of  $\{1, \dots, M\}$  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 sub-areas  $b \in \{1, \dots, B\} = \mathcal{B}$  and assume all consumers residing in a sub-area are located at the sub area's centroid. (In our study, we use tracts, as defined by the U.S. Census, with an average of 1.2 square miles in Cook county.) Thus, all potential consumers residing in a given sub-area  $b \in \mathcal{B}$  and belonging to a

given socio-economic group  $m \in M$ , experience the same *mean* utility value for all outlets as well as the no-purchase option.

Assuming the distributions of the random noise terms,  $\{\eta_{ij} : j = 1, \dots, J\}$ , are Gumbel (or doubly exponential), with common scale parameter  $\mu$ , 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:

$$S_{j,b,m}(P, W, X|\beta, \delta, \gamma, \xi, \pi) = \frac{e^{(X'_j\beta - D_{bj}\delta - P_j\gamma - W_j\xi)/\mu}}{e^{(\pi M + \beta_0)/\mu} + \sum_{t=1}^J e^{(X'_t\beta - D_{bt}\delta - P_t\gamma - W_t\xi)/\mu}}, j = 1, \dots, J; b = 1, \dots, B; m = 1, \dots, M. \quad (3)$$

Without loss of generality and to simplify the notation we may choose the units in which the utility levels are measured such that the scale parameter  $\mu = 1$ . Also, as is evident from (3), the market shares are invariant to a common additive shift in the utility measures. Therefore, again without loss of generality, we may specify the choice of the parameters  $\beta$  such that

$$\max \{X'\beta : X' = X'_j \text{ for some } j = 1, \dots, J\} = 1 \quad (4)$$

(For example, in our study, the  $X$ -vector is a unit vector which identifies the chain to which the outlet belongs: thus (4) simplifies to  $\beta_{max} \equiv \max_{j=1, \dots, J} \beta_j = 1$ , a constraint which is easily incorporated into the optimization problem that is solved in our (GMM)-estimation method, see §5.)

Multiplying the market shares with the number of consumers in each geographic region  $b$  and demographic group  $m$  allows us to specify expected aggregate sales in an outlet as a function of the various parameters  $\theta \equiv \{\beta, \delta, \gamma, \xi, \pi\}$  in the utility equations:

$$Q_j(P, W, X|\beta, \delta, \gamma, \pi) = \sum_b \sum_m h(b, m) S_{j,b,m}(P, W, X|\beta, \delta, \gamma, \pi, ) \quad (5)$$

Here  $h(b, m)$  denotes the number of residents (or commuters) in sub-area  $b$  who belong to demographic group  $m$ .

### 3.2. The outlets' cost structure

We assume the outlets' cost structure 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:

$$C_j(Q_j) = (c_{k(j)} + \epsilon_j)Q_j + [d_{k(j)} + u_j]/W_j, j = 1, \dots, J \quad (6)$$

Here

$k(j)$  = the index of the chain to which outlet  $j$  belongs,  $j = 1, \dots, J$

$c_k$  = the average variable food, labor and equipment cost rate per customer for an outlet of chain  $k, k = 1, \dots, K$

$d_k$  = the average variable capacity cost rate for an outlet of chain  $k, k = 1, \dots, K$

$\epsilon_j$  = a noise term, denoting the difference between outlet  $j$ 's variable cost rate and the norm for his chain,  $j = 1, \dots, J$

$u_j$  = a noise term, denoting the difference between outlet  $j$ 's variable capacity cost rate and the norm for his chain,

$j = 1, \dots, J$

Each outlet's marginal cost rate as well as the capacity cost rate are equal to a chain-specific cost plus a zero-mean unobserved component. This specification is supported by the franchisers' effort to create a uniform customer experience across their outlets, by standardizing the equipment, preparation process, and food components used at each of its outlets. The unobserved shock to the cost rate comes from outlet specific conditions. For example, a smaller kitchen could create crowding and reduce efficiency.

The affine cost structure in (6) arises in several queueing models which may describe the service process of an outlet. For example, the structure in (6) 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 (2006) have shown that the cost structure in (6) applies to a general Jackson network, assuming the variable capacity costs are proportional with the service rates installed at the various nodes of the network. Thirdly, 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 (6) arises as a close approximation, see Allon and Federgruen (2006). This approximation is

based on so-called exponential approximations for the tail probability of steady-state delays, also referred to as Cramer-Lundberg approximations. The exponential approximation states the existence of constants  $a, \xi > 0$  such that  $Pr[D > x] \approx ae^{-\xi x}$ , i.e.  $\lim_{x \rightarrow \infty} e^{\xi x} Pr[D > x] = 0$ . 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 (2006) for additional queueing models resulting in affine cost structures of type (6). Allon and Federgruen (2006) also show that an even larger set of queueing models give rise to cost functions of the type  $C_j(Q_j) = (c_{k(j)} + \epsilon_j)Q_j + \sqrt{(d_{k(j)}^{(1)} + u_j^{(1)})Q_j^2 + (d_{k(j)}^{(2)} + u_j^{(2)})Q_j/W_j + (d_{k(j)}^{(3)} + u_j^{(3)})/W_j^2}$  for parameters  $\{d_k^{(1)}, d_k^{(2)}, d_k^{(3)} : k \equiv 1, \dots, K\}$  and mean zero noise terms  $\{u_j^{(1)}, u_j^{(2)}, u_j^{(3)}\}$ . Our estimation method can be adapted to this more general cost structure.

### 3.3. The Price Competition Model

We are now ready to analyze the price competition model which 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, an issue centrally considered in Thomadsen (2005a). In view of (6)

$$\pi_j(P, W, X, \theta) = (P_j - c_{k(j)} - \epsilon_j)Q_j(P, W, X|\theta) - [d_{k(j)} + u_j]/W_j, \quad j = 1, \dots, J \quad (7)$$

denotes firms  $j$ 's profit level as a function of all prices charged by the various outlets. The following theorem shows that the price competition model has a unique equilibrium, which is the unique solution of the system of equations (8), below. The theorem also shows that the equilibrium can be computed effectively by a tâtonnement scheme. This scheme iteratively identifies each firm's best response to the prices of the competing firms, see Topkis (1998) for details. The availability of a simple computational scheme, under estimated parameter values, is important when conducting the counterfactual study in Section 6.2. We refer to Allon et al. (2009) for a proof of the following theorem.

**Theorem 3.1(a)** *The price competition model has a unique equilibrium in the interior of the price cube  $\times_{j=1}^J [c_{k(j)} + \epsilon_j, c_{k(j)} + \epsilon_j + 1/\gamma]$ . The equilibrium is the unique solution of the system of equations:*

$$Q_j(P, W, X|\theta) + (P_j - c_{k(j)} - \epsilon_j) \frac{\partial Q_j(P, W, X|\theta)}{\partial P_j} = 0, \quad j = 1, \dots, J \quad (8)$$

(b) *Fix an outlet  $j = 1, \dots, J$ . Let  $P_{-j}$  denote the vector of prices charged by the other outlets and let  $P_j^*(P_{-j})$  denote outlet  $j$ 's best response to the prices charged by the competing outlets. The best response function  $P_j^*(\cdot)$  is increasing.*

(c) *For a given parameter vector  $\theta$ , the unique equilibrium can be found by a simple tâtonnement scheme, starting with the lower-bound price vector  $p^{min} \equiv [c_{k(j)} + \epsilon_j : j = 1, \dots, J]$ .*

The following is a simple expression for the partial derivatives of the firms' sales quantities with respect to price:

$$\frac{\partial Q_j(P, W, X|\theta)}{\partial P_j} = -\gamma \sum_{b=1}^B \sum_{m=1}^M h(b, m) \left( 1 - \frac{S_{j,b,m}(P, W, X|\theta)}{h(b, m)} \right) S_{j,b,m}(P, W, X|\theta) \quad (9)$$

All three parts of Theorem 3.1 are noteworthy, as the results apply to general price competition models with so-called multinomial logit demand function with random coefficients, and any affine cost structure<sup>3</sup>. Thomadsen (2005b), recently, identified some sufficient conditions for the existence of an equilibrium but concludes his paper, stating "While the theorem provides a limited set of conditions under which there exists a pure strategy equilibrium . . . , and while existence is shown for any distribution of consumers and firms, what is lacking are proofs of uniqueness".

In matrix notation, the equilibrium conditions (8) can be stated as:

$$Q(P, X, W) + \Omega(P - c - \epsilon) = 0 \quad (10)$$

where  $\Omega$  is a diagonal  $J \times J$  matrix whose the  $j$ -th diagonal element  $\Omega_{j,j} = \frac{\partial Q_j}{\partial P_j}$ . Essential to our estimation method is that the vector of cost rate residuals  $\epsilon$  therefore satisfies:

$$\epsilon = P - C + \Omega(P, X, W|\theta)^{-1} Q(P, X, W|\theta), \quad (11)$$

<sup>3</sup> For example, in their classical paper Berry et al. (1995) note: "We assume that a Nash equilibrium to this pricing game exists, and that the equilibrium prices are in the interior of the firms' strategy sets (the positive orthant). While Caplin and Nalebuff (1991) provide a set of conditions for the existence of an equilibrium for related models of single product firms, their theorems do not easily generalize to the multi product case. However, we are able to check numerically whether our final estimates are consistent with the existence of an equilibrium. Note that none of the properties of the estimates require uniqueness of equilibrium, although without uniqueness it is not clear how to use our estimates to examine the effects of policy and environmental changes."

with  $\theta' = (\beta', \gamma', \delta', \pi', \eta')$ , and identity which follows from (10).

Finally, when multiple outlets are owned by the same franchisee, and the total number of independent franchisees is  $F < J$ , the first order conditions (8) can also be written in the form (11), however with a non-diagonal matrix  $\Omega$ . See Thomadsen (2005b) for details.

### 3.4. The Two-Stage Competition Model

In Section 6.2, we investigate how the competing firms in the industry select their service levels, i.e. their waiting time standards. As explained at the beginning of this section, the waiting time standards are determined at the chain level in advance of the outlets' price choices. We therefore assume that the industry's waiting time standards represent the equilibrium in a two-stage game, with the above price competition model as the second stage, and a first stage game in which the chains act as the "players".

In the fast food industry, the focus of our study, it has been widely reported (see, for example, Thomadsen (2005a) and Lafontaine and Shaw (1999)) that the franchisee pays a fixed, periodic, franchise fee along with a percentage of the revenues. However, it is unclear how these two parameters are determined as a function of the desired waiting time standard. We therefore assume that each chain attempts to maximize aggregate profits in its supply chain, i.e. its own profits as well as those of its franchisees. (The aggregate profit measure is independent of the transfer payment scheme.)

It remains an open question whether a pure strategy Nash equilibrium exists in the two-stage game and whether it is unique. However, when searching for an equilibrium we employ the tâtonnement scheme, described in subsection 3.3. To evaluate the profit consequences of any vector of waiting times standard, we solve the associated price competition model, which has a unique equilibrium, see Theorem 3.1(a).

## 4. Data

We have studied the fast-food industry in Cook County, Illinois. We have chosen 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 Magazine 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 adopted to improve speed-of-service include timer systems that allow in-store managers

as well as regional and national offices to monitor waiting times at outlets, as well as the outsourcing of drive-thru order taking. There is a plethora of anecdotal evidence that the industry is reacting to consumer expectations. The same 2007 QSR Magazine Drive-Thru study reported that 70% of surveyed customers said speed is an important factor in the drive-thru experience. The 2008 study reports that this trend is continuing, with the fastest chain, Wendy's, shaving off an additional 7 seconds from the average waiting time in the previous year.

We use as our data set, all fast-food outlets belonging to chains selling hamburgers in Cook County Illinois. 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. We consider only chains with a presence of more than 5 outlets in the county. This results in a total 388 outlets belonging to McDonald's (173), Burger King (92), Wendy's (62), White Castle (42), Dairy Queen (10), and Steak 'n Shake (9) chains.

For each outlet we gathered prices for the franchise's signature burger, a small fries order, and a small soft-drink by calling the location in 2006. 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. These prices have been aggregated to compute the price of a complete standard meal at that outlet. It is worth noting that we tried to call the chains themselves for price information or for any pricing guidelines that they may give to their franchisees. We were told that even the practice of suggesting prices to the outlets is illegal (see the discussion at the beginning of §3). Thus, waiting times are selected centrally by the chains but prices are chosen by the individual outlets. Indeed, we have noticed 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 in the county, see Table 3. This supports the two stage competition model we have postulated in the previous section.

As mentioned, 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 chain outlet they go to. 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 subsection 3.1. To arrive at the average waiting time standards for the different chains, we have employed the QSR Magazine's 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 below. 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]. We have performed a two-

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

Rank	Chain	Mean Wait (sec)
2	WENDY'S	173.34
4	BURGER KING	192.29
14	MCDONALD'S	224.27
15	DAIRY QUEEN	231.85
21	STEAK'N SHAKE	264.3
24	WHITE CASTLE	269.45

sample t-Test assuming unequal variances on all the national waiting time observations for the six chains with a major presence in Cook County. We have verified that waiting time observations for different chains were indeed drawn from different distributions, (with the exception of the McDonald's and Dairy Queen, and the White Castle and Steak 'n Shake pairs; note from Table 1, that the mean waiting times are virtually identical within each pair 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 2. 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 the table.

Demographic and geographic information was gathered with a very fine granularity, 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 1343 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

**Table 2 Two-Sample t-Test on National Chain-Wide Wait Time Observations**

Chain	Num. Obs.	Mean Wait (sec)	Std. Dev.	Mc-Donald'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	–

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 black and white consumers only because these are the racial groups for which we had the necessary data for employing the macro moments discussed in Section 6. As mentioned in Section 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 broken down by age group, race, and gender. 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). The Bureau of Transportation Statistics data are not broken down by age, gender and race combination at the tract level in Cook County, so, we estimated the population numbers of each combination by assuming the three social attributes are independent. This means that 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, i.e. lunch, while at work, and another meal, i.e. dinner, while at home. Distinguishing between commuters and residents, two genders and two racial groups, as well as among 5 age brackets, we have thus divided the population into 40 different socio-economic groups. The distance from the consumer to each outlet is calculated as the distance from the tract centroid in which the consumer is located. To compute the distances between the various tract centroids and the restaurant locations, we have 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. These are outlet specific variables that are correlated with one or more of the independent variables but not with the noise terms  $\{\epsilon_j : j = 1, \dots, J\}$  in the cost rates. Following the recommendation in Thomadsen (2005a), we have selected the following instrumental variables:  $V_{1j}$  = the distance from outlet  $j$  to the nearest outlet,  $V_{2j}$  = the number of outlets within two miles of outlet  $j$ ,  $V_{3j}$  = the population density in the tract to which outlet  $j$  belongs, and  $V_{4j}$  = the worker density in this tract. Table 3 shows summary statistics for these as well as the price variables.

**Table 3** Summary Statistics for Collected 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 (100K/sq mi)	0.09	0.09	1.71E-04	0.80
Worker Density (100K/sq mi)	0.04	0.05	1.33E-03	0.36

## 5. Estimation

As mentioned in the introduction, the major hurdle to our estimation of 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. The unavailability of sales data prevents the use of standard regression-type estimation techniques. Instead, we employ a technique that estimates the parameters on the basis of the equilibrium conditions, specified in subsection 3.3, which characterize the unique Nash equilibrium in the second-stage price competition among the outlets, see Theorem 3.1. In the context of industrial organization studies, this so-called Generalized Method of Moments (GMM) technique for circumventing a lack of demand data, was first introduced by Berry et al. (1995). See Nevo (2000) for a clear exposition.

The equilibrium conditions (11) represent a system of equations which involve only the *observed* price vector  $P$ , waiting time standards  $W$ , outlet attribute matrix  $X$ , and distances  $\{D_{j,b} : j = 1, \dots, J, b =$

$1, \dots, B\}$ , as well as the *unknown* parameter string. (In particular, the system of equations does not involve the unobservable sales volumes.) Because of the endogeneity of the price vector  $P$ , these variables are correlated with the cost rate noise terms  $\{\epsilon_j : j = 1, \dots, J\}$ . This prevents the usage of standard maximum likelihood estimation techniques. In addition, the latter requires us to postulate specific distributions for the noise terms  $\epsilon$ .

The (GMM) technique overcomes both difficulties. It employs a vector of so-called instrument variables  $Z_j \equiv \{Z_{1j}, \dots, Z_{rj}\}$  which are correlated with (some of) the explanatory variables  $\{P, X, W, D\}$ , but uncorrelated with the cost rate noise terms  $\epsilon$ , i.e.

$$E[Z_{lj}\epsilon_j] = 0 \text{ for all cost rates } l = 1, \dots, r \text{ and all outlets } j = 1, \dots, J \quad (12)$$

Our instruments are based on the four instrumental variables  $V_{1j}, V_{2j}, V_{3j}, V_{4j}$  defined in Section 4. In order 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.e. the columns of the matrix  $X$ , to arrive at a total of 24 instruments: for all  $j = 1, \dots, J$ ,  $Z_j \equiv \{Z_{l,k,j} = V_{lj} \cdot X_{kj}, l = 1, \dots, 4, k = 1, \dots, K\}$ . 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 any cost rate differentiated outlets  $j$  is experiencing vis-a-vis the chain norm. In other words, it is reasonable to assume that the moment conditions (12) apply. In view of the population moment conditions (12) we must have that for the proper parameter vector  $\theta$ , the sample average of the vector of random variables  $\{Z_j^t \epsilon_j, j = 1, \dots, J\}$

$$G(\theta) = \frac{1}{J} \sum_{j=1}^J Z_j^t \epsilon_j(\theta) \quad (13)$$

is as close to zero as possible. The (GMM) estimator thus determines a parameter vector  $\hat{\theta}$  which minimizes a quadratic function of this sample average. More specifically the (GMM) estimate is the vector  $\hat{\theta}$  which optimizes

$$\min_{\theta} G(\theta)' A G(\theta) \quad (14)$$

where  $A$  is a weighting matrix for the moments.

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, as this matrix is not available, a-priori, we follow the commonly used two-step estimation procedure where the first step is used to estimate the asymptotic variance-covariance matrix. In the first step, we use the GMM with the pre-specified weighting matrix  $A_1 = I$  to get a consistent initial estimator  $\hat{\theta}_1$ . We use  $\hat{\theta}_1$  to estimate the asymptotic variance-covariance matrix of the moment conditions,  $A_2 = (E[G(\hat{\theta}_1)G(\hat{\theta}_1)'])^{-1}$ . We then run the GMM procedure a second time with this new weighting matrix to arrive at our parameter estimate  $\theta_2$ . For a nice discussion of the two step method and a proof that the inverse of the asymptotic variance -covariance matrix is the optimal weighting matrix, see Hall (2005).

There are well documented technical difficulties associated with the optimization problem (14). Its objective function has many local optima quite far from the globally optimal solution. In addition, there are large regions where this function is close to flat, creating formidable difficulties for standard gradient methods.

To mitigate these difficulties, we restrict the feasible region for the parameter vector  $\theta$  by imposing several reasonable constraints. Let

$$\begin{aligned} con(\theta)_{b,m} &= \sum_{j=1}^J S_{j,b,m}(P, W, X|\theta)/h(b, m), b = 1, \dots, B; m = 1, \dots, M \\ &= \text{the fraction of the population in tract } b \text{ and socio-economic group } m \text{ which purchases a fast food meal;} \end{aligned}$$

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

$\underline{con}$  = a lower bound for the fraction of the population in any geographical area and any socio-economic group to purchase a fast food meal;

$$\begin{aligned} \hat{c}_k(\theta) &= \text{best estimate of chain } k\text{'s standard cost rate } c_k \\ &= \frac{1}{J_k} \sum_{j \in J_k} [P_j + \Omega(P, X, W|\theta)^{-1} Q(P, X, W|\theta_j)], k = 1 \dots, K, \text{ where} \\ J_k &= \{j : k(j) = k\} \text{ denotes the set of outlets belonging to chain } k \end{aligned}$$

$$c_{k,min} = 0$$

$$c_{k,max} = \min_{j \in J_k} P_j : k = 1, \dots, K.$$

We impose the constraints:

$$\underline{con} \leq con(\theta)_{b,m} \leq \overline{con}, \text{ for all } b = 1 \dots, B \text{ and } m = 1 \dots, M \quad (15)$$

$$c_{k,min} \leq \hat{c}_k(\theta) \leq c_{k,max}, \text{ for all } k = 1 \dots, K \quad (16)$$

Thus, instead of solving the unconstrained optimization problem (13), we propose solving the constrained optimization problem:

$$(P) \quad \min_{\theta} \{(12) \text{ s.t. (15) and (16)}\}.$$

To solve the constrained optimization problem, we propose replacing the soft constraints (15) and (16) by penalty functions added to the objective function. We thus obtain:

$$G'(\theta)AG(\theta) + \Lambda \sum_b \sum_m \{\log[\overline{con} - con(\theta)_{b,m}] + \log[con(\theta)_{b,m} - \underline{con}]\} \quad (17) \\ + \sum_k \log[c_{k,max} - \hat{c}_k(\theta)] + \log[\hat{c}_k(\theta) - c_{k,min}]$$

with  $\Lambda$  as the weighting factor, a multiplier applied to the penalty functions.

The fact that the chain indicator variables and the waiting time standard are common to all outlets in a chain creates a co linearity problem. We therefore normalize one of the chain indicator  $\beta$ -parameters (that of Wendy's) to zero. (Recall that all market share formula in Section 3.1 are invariant to a common additive shift in the utility measures.)

We have developed a special algorithm to solve (P) via the modified objective (17). The algorithm begins with a large value for  $\Lambda$ , the weight of the penalty functions, relative to the objective function value at the starting point and involves 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  $\Lambda$  is halved and the quasi-Newton search re-run. In the first iteration, when the penalty  $\Lambda$  is large, this results in the algorithm moving to a point which 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 (14).

To arrive at the reported estimates, we used a process in which, in the first stage, we took 19 starting points and ran both the above algorithm as well as the general KNITRO algorithm with default options - resulting in 38 estimates. We then took the (almost) 20 best local optima found, for the objective function (13), generated weighting matrices for each of these points and, in the second stage, ran our algorithm as well as the KNITRO algorithm from each of these 20 points, generating 40 final estimates.

In order to validate the statistical significance of these estimates, we constructed confidence intervals using a bootstrapping procedure. This procedure is used when attempting to measure the variance of an estimator when there is no sample data available beyond those used to obtain the estimate. The idea is to use subsets of the sample and calculate the value of the estimator on each subset in order to estimate the variance. To that end, we selected 20 random subsets of the tracts, and ran the algorithm on each subset for each of the 20 (second stage) starting points of the two algorithms, resulting in a total of 800 parameter values. Each subset has 134 tracts (10% of total number of tracts). We estimated the coefficients for each sub-model and used the empirical distribution to construct the confidence intervals for each parameter.

## 6. Results

In this section, we report the results of the estimation process. We focus on the key parameters of interest, emphasizing those that turned out to be statically significant. Table 4 reports the estimated value of each of the main demand coefficients: price, waiting time and distance sensitivity ( $\gamma, \delta, \xi$ ). We also report the estimated marginal cost for each of the top three chains: McDonald's, Burger King and Wendy's. The table reports, in addition, the 90% percent confidence intervals for each of the reported estimates.

**Table 4** Estimates of Price, Waiting Time and Distance Coefficients

Variable Name	Coefficient Est.	90% Confidence Interval
Price Sensitivity (\$)	0.4364	[0.3349, 0.6088]
Waiting Time Sensitivity (sec)	0.0259	[0.0070, 0.0541]
Distance Sensitivity (mi)	0.00350	[0.0001, 0.0143]
Marginal Cost McD	2.563	[1.972, 3.313 ]
Marginal Cost BK	2.460	[1.868, 3.209]
Marginal Cost WN	2.354	[1.746, 3.105]

First note that both the price sensitivity and waiting time sensitivity parameters  $\gamma$  and  $\xi$ , see (1), are significantly positive, at the 90% confidence level. (The two are, in fact, significant at the 99% confidence

level). We thus conclude that both price and waiting time parameters have a significant impact on the consumer's decision at which outlet to purchase a fast-food meal, if at all. This result confirms 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 addition, our estimates in Table 4, indicate that consumers attribute a very high cost to the time they spend waiting. In particular, to overcome an additional *second* of waiting time, an outlet will need to compensate an average customer by as much as \$0.06 in a meal whose typical price ranges from \$2.25 to \$6. This corresponds to an hourly cost rate of more than *ten* times the (pre-tax) average wage of \$18/hour. Even when considering the 90% confidence intervals for the estimates, comparing (opposite) extreme values of these confidence intervals, reveals that the average consumer assigns a cost to waiting, which corresponds with an hourly rate of at least \$40, more than twice the average pre-tax wage in the US. Since price differences in this, as in many other industries, are rather modest, this implies that in the drive-thru market waiting time plays a more significant role than pricing in explaining sales volumes. Moreover, these results seem to justify the continuing substantial trend of investments chains make to improve their waiting time performance.

We also observe that the sensitivity to the distance measure is not statistically significant. The contrast with the waiting time sensitivity is, at first, striking. After all, the consumer's need to traverse a given distance translates into an amount of time expended in traveling as well as the waiting process in the drive-thru queue. The fact that this loss of time is valued so differently from the time spent waiting in the drive-thru queue itself is consistent with finding in the behavioral economics literature, see e.g. Kahneman and Tversky (1984) and Larson (1987) which reveal that individuals value time very differently, depending on the context and the degree to which the time is spent is pleasurable or not: most people mind time spent driving far less than time waiting idly; many even enjoy the ride. In addition, the distance between the consumers' residence and the outlet 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, and thus there is no real disutility to distance. A last explanation is the fact that our study was carried out in Cook County, which is fairly populated and dense in fast food outlets.

An outlet derives a shift of its utility value if it belongs to either the McDonald's and Burger King chains: the shift is positive, yet not statistically significant. On the other hand a Dairy Queen franchise has a negative shift in the outlet's mean utility value, which is again insignificant. Due to the lack of statistical significance, these  $\beta$ -parameters are not reported in the paper, see (1).

As stated above, one of the advantages of the estimation method is that it also allows us to estimate the marginal costs. The average marginal costs for the three largest chains: McDonald's, Burger King and Wendy's are fairly close. These, in conjunction with the average prices can be used to infer market power of each of the chains. Given that the average prices for these chains are \$4.96, \$4.86 and \$4.75, we get that the markups are \$2.414, \$2.407 and \$2.413. One can infer that the market power of all chains is quite significant, and note that the differences between them are insignificant.

### 6.1. 1 Stage Estimation With Macro Moments

In order to test the robustness of our estimates and also to attempt to improve the efficiency of our estimates, we supplemented the twenty-four micro-moments, introduced in §5, with additional so-called macro-moments. Imbens and Lancaster (1994) suggest supplementing micro-moments with macro-moments to increase the efficiency of the estimates, and this approach has been used in industrial organization studies by Petrin (2002) and Davis (2006).

We have added macro-moments that are based on 3 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. The following twelve macro-moments were added to the micro-moments (10 based on comparisons between age brackets, one between genders and one between races).

$$G_{0-9,10-16}(\theta) = \frac{1}{J} \sum_{j=1}^J \left[ R_{0-9} \frac{Q_{j,10-19}(\theta)}{Pop_{10-19}} - R_{10-19} \frac{Q_{j,0-9}(\theta)}{Pop_{0-9}} \right] \quad (18)$$

...

$$G_{40-59,60+}(\theta) = \frac{1}{J} \sum_{j=1}^J \left[ R_{40-59} \frac{Q_{j,60+}(\theta)}{Pop_{60+}} - R_{60+} \frac{Q_{j,40-59}(\theta)}{Pop_{40-59}} \right] \quad (19)$$

$$G_J(\theta) = \frac{1}{J} \sum_{j=1}^J \left[ R_{Male} \frac{Q_{j,Female}(\theta)}{Pop_{Female}} - R_{Female} \frac{Q_{j,Male}(\theta)}{Pop_{Male}} \right] \quad (20)$$

$$G_J(\theta) = \frac{1}{J} \sum_{j=1}^J [R_{Black} \frac{Q_{j,White}(\theta)}{Pop_{White}} - R_{White} \frac{Q_{j,Black}(\theta)}{Pop_{Black}}] \quad (21)$$

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. As suggested in Thomadsen (2005a) The macro-moments 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. The addition of these moments yielded the results detailed in Table 5, which reports the estimated coefficients for the main parameters.

**Table 5 Estimates of Price, Waiting Time and Distance Coefficients Employing Additional Macro-Moments**

Variable Name	Coefficient Est.	Confidence Interval
Price Sensitivity(\$)	0.4411	[0.267, 0.6098]
Waiting Time Sensitivity (sec)	0.025	[0.0016 0.0490]
Distance Disutility (mi)	0.0455	[5.7E-05 1.409 ]

The results are reported based on a single stage of estimation. We were unable to perform the second estimation stage since the moment variance-covariance matrices at most candidate optima were close to singular. The estimates are close to those obtained without the macro-moments, thus showing that our results are robust. Due to the fact that we could not perform the second stage of the estimation procedure, the new confidence intervals are slightly wider than when estimating without the macro-moments.

## 6.2. Counter Factuals

How much, then, is it worth to reduce the waiting time standards? We mentioned the industry maxim that a 7 second reduction in waiting times increase a chain's market share by 1%. We investigated the impact of a *single* chain reducing its waiting time standard by 7 seconds, allowing all outlets to adjust their prices to the new price equilibrium. The results of this experiment can be seen in Table 6. In the first two rows, we give the estimated daily demand and market share of each chain at the current waiting time standards and prices. The subsequent rows show the resulting changes in the chains' market share and demand volume.

The percentage of the total market captured at the current waiting time standards closely matches the results in 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 fastest chain 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.) Even more importantly, an unmatched reduction of McDonald’s waiting time standard by 7 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. Any chain’s unilateral waiting time reduction,

**Table 6 Change In Market Share Following 7 second Wait Standard Reduction at a Chain**

	McD	BK	WN	Wh. Castle	DQ	S 'n S	% of Total Market
Initial Demand	1.15E+06	2.22E+05	1.11E+05	2.27E+04	2.73E+03	4.84E+03	22.92%
Init. Market Share	75.98%	14.68%	7.33%	1.5%	0.18%	0.32%	
McD ( $\Delta$ Demand)	1.82E+05	-7.44E+03	-3.71E+03	-761.76	-91.38	-162.07	25.50%
( $\Delta$ Market Sh.)	3.16%	-1.93%	-0.96%	-0.20%	-0.02%	-0.04%	
BK	-7.64E+03	4.24E+04	-737.17	-151.31	-18.15	-32.19	23.43%
	-2.16%	2.42%	-0.21%	-0.04%	-0.01%	-0.01%	
WN	-3.83E+03	-739.64	2.16E+04	-75.78	-9.09	-16.12	23.18%
	-1.09%	-0.21%	1.33%	-0.02%	0.00%	0.00%	
WC	-787.63	-152.22	-75.98	4.51E+03	-1.87	-3.32	22.97%
	-0.23%	-0.04%	-0.02%	0.29%	0.00%	0.00%	
DQ	-94.54	-18.27	-9.12	-1.87	542.63	-0.40	22.93%
	-0.03%	-0.01%	0.00%	0.00%	0.04%	0.00%	
SS	-167.66	-32.40	-16.17	-3.32	-0.40b	962.04	22.93%
	-0.05%	-0.01%	0.00%	0.00%	0.00%	0.06%	

is likely to result in waiting time changes by the competing firms. Indeed, between 2005 and 2008, almost all chain have gradually reduced their waiting time standards, with McDonald’s going from 224 to 158 seconds and Wendy’s reducing its standard from 173 to 131. To investigate the chains’ equilibrium waiting time choices, we have considered the two-stage game outlined in Section 3.4. In the first stage, chains choose the target waiting times and in the second stage the outlets compete in prices, taking the waiting

times as given. As explained in Section 3.4, we assume that each chain maximizes chain wide profits. This counterfactual study requires knowledge of the chains' variable cost ratios  $\{d_k\}$ , see (6). Like the  $\{c_k\}$ -parameters these are not directly observable; moreover they cannot be estimated either, since they do not appear in the equilibrium condition (8). All we know is that  $0 \leq d_k \leq c_k$ . We have, therefore, computed the two-stage equilibrium assuming  $d_k = 0.5c_k, k = 1, \dots, K$  and a minimum feasible waiting time standard of 2 minutes, approximately 10% lower than the 2008 best practice of 131 seconds (by Wendy's). Indeed, in equilibrium all chains reduce their waiting time standards to the minimum feasible value, providing an analytical justification for the industry trend over the past 5-10 years.

In conclusion, reducing waiting time standards pays off handsomely in the fast food industry. Consumers assign an implicit value to waiting in the drive-thru queue which amounts to at least \$40/hr, more than twice the pre-tax U.S. wage. A 7 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 but for a large chain like McDonald's, it would result in an increase by more than 3% and an increase in its sales volume by 15%. The competitive dynamics are such that, to the extent feasible via marginal 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.

## References

- Allon, G., A. Federgruen. 2006. Service competition with general queueing facilities. *Forthcoming in Operations Research*.
- Allon, G., A. Federgruen. 2007. Competition in service industries. *Operations Research* **55**(1).
- 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*.
- Armony, M., M. Haviv. 2001. Price and delay competition between two service providers Technical Report, Stern School, NYU, NY.
- Berry, S., J. Levinsohn, A. Pakes. 1995. Automobile prices in market equilibrium. *Econometrica* **63**(4) 841–890.
- Berry, S.T. 1994. Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics* 242–262.
- Bresnahan, T.F. 1987. Competition and collusion in the American automobile industry: The 1955 price war. *The Journal of Industrial Economics* 457–482.

- Cachon, G.P., F. Zhang. 2007. Obtaining fast service in a queueing system via performance-based allocation of demand. *Management Science* **53**(3) 408.
- Cachon, P. C., P. T. Harker. 2002. Competition and outsourcing with scale economies. *Management Science* **48**(10) 1314–1333.
- Caplin, A., B. Nalebuff. 1991. Aggregation and imperfect competition: On the existence of equilibrium. *Econometrica: Journal of the Econometric Society* 25–59.
- Chen, H., M.Z. Frank, O.Q. Wu. 2005. What actually happened to the inventories of American companies between 1981 and 2000? *Management Science* **51**(7) 1015.
- Chen, H., M.Z. Frank, O.Q. Wu. 2007. US Retail and Wholesale Inventory Performance from 1981 to 2004. *Manufacturing & Service Operations Management* **9**(4) 430.
- Chen, H., Y-W. Wan. 2003. Price competition of make-to order firms. *IIE Transactions* **35**(9) 817–832.
- Davis, M.M., T.E. Vollmann. 1990. A framework for relating waiting time and customer satisfaction in a service operation. *Journal of Services Marketing* **4**(1) 61–69.
- Davis, P. 2006. Spatial competition in retail markets: movie theaters. *RAND Journal of Economics* 964–982.
- 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. *The Review of Economics and Statistics* 149–153.
- De Vany, A.S., T.R. Saving. 1983. The economics of quality. *The Journal of Political Economy* **91**(6) 979–1000.
- Deacon, R.T., J. Sonstelie. 1985. Rationing by waiting and the value of time: Results from a natural experiment. *The Journal of Political Economy* **93**(4) 627–647.
- Dube, J.P.H., J.T. Fox, C.L. Su. 2008. Improving the Numerical Performance of BLP Static and Dynamic Discrete Choice Random Coefficients Demand Estimation. *Working Paper. University of Chicago, Chicago, IL* .
- Feenstra, R., J. Levinsohn. 1995. Estimating markups and market conduct with multidimensional product attributes. *Review of Economic Studies* 19–52.
- Gans, N. 2002. Customer loyalty and supplier quality competition. *Management Science* 207–221.
- Gans, N., G. Knox, R. Croson. 2007. Simple models of discrete choice and their performance in bandit experiments. *Manufacturing & Service Operations Management* **9**(4) 383.
- Gans, N., G. Koole, A. Mandelbaum. 2003. Telephone call centers: Tutorial, review and research prospects. *Manufacturing & Service Operations Management* **5**(2) 79–141.
- Garber, Amy. 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* .
- Gaur, V., M.L. Fisher, A. Raman. 2005. An econometric analysis of inventory turnover performance in retail services. *Management Science* **51**(2) 181–194.
- Gilbert, S.M., Z.K. Weng. 1998. Incentive effects favor nonconsolidating queues in a service system: The principal-agent perspective. *Management Science* 1662–1669.

- Hall, A.R. 2005. *Generalized method of moments*. Oxford University Press, USA.
- Hall, J., E. Porteus. 2000. Customer service competition in capacitated systems. *Manufacturing & Service Operations Management* **2**(2) 144.
- Hasija, S., E.J. Pinker, R.A. Shumsky. 2007. Call center outsourcing contracts under information asymmetry. *Management Science* .
- Hassin, R., M. Haviv. 2003. *To Queue or Not to Queue: Equilibrium Behavior in Queueing Systems*. Kluwer Academic Publishers, Boston, MA.
- Hess, S., M. Bierlaire, J.W. Polak. 2005. Estimation of value of travel-time savings using mixed logit models. *Transportation Research Part A* **39**(2-3) 221–236.
- Hughlett, M. Nov 28, 2008. Drive-throughs done right ring up returns. *Chicago Tribune* .
- Imbens, G.W., T. Lancaster. 1994. Combining micro and macro data in microeconomic models. *The Review of Economic Studies* 655–680.
- Jargon, Julie. 2006. McD's service stalls at drive-thru. *Crain's Chicago Business* .
- Kahneman, D., A. Tversky. 1984. Choices, values, and frames. *American Psychologist* **39**(4) 341–50.
- Kalai, E., M. Kamien, M. Rubinovitch. 1992. Optimal service speeds in a competitive environment. *Management Science* **38**(8) 1154–1163.
- Lafontaine, F., K.L. Shaw. 1999. The dynamics of franchise contracting: Evidence from panel data. *Journal of Political Economy* **107**(5) 1041–1080.
- Larson, R.C. 1987. Perspectives on queues: social justice and the psychology of queueing. *Operations Research* 895–905.
- Lederer, P. J., L. Li. 1997. Pricing, production, scheduling and delivery-time competition. *Oper. Res.* **45**(3) 407–420.
- Levhari, D., I. Luski. 1978. Duopoly pricing and waiting lines. *European Economic Review* **11** 17–35.
- Li, L., Y. S. Lee. 1994. Pricing and delivery-time performance in a competitive environment. *Management Science* **40**(5) 633–646.
- Loch, C. 1991. Pricing in markets sensitive to delays. *Ph.D. Dissertation, Stanford University, Stanford, CA* .
- Luski, I. 1976. On partial equilibrium in a queueing system with two servers. *The Review of Economic Studies* **43** 519–525.
- Mueller, C.D. 1985. Waiting for Physicians' Services: Model and Evidence. *The Journal of Business* **58**(2) 173–190.
- Musalem, A., M. Olivares, E. Bradlow, C. Terwiesch, D. Corsten, U. Hall, S. Srinivasan, K. Pauwels. 2009. Structural estimation of the effect of out-of-stocks. *Working Paper* .
- Nevo, A. 2000. A practitioners guide to estimation of random coefficients logit models of demand. *Economics & Management Strategy* **9**(9) 513–548.

- Nuckolls, M. W. 2007. Building a better drive-thru: Consumer drive-thru preferences. *QSR Magazine* **Drive-Thru Time Study**.
- Olivares, M., G. Cachon. 2007. Competing Retailers and Inventory: An Empirical Investigation of US Automobile Dealerships .
- Olivares, M., C. Terwiesch, L. Cassorla. 2008. Structural estimation of the newsvendor model: an application to reserving operating room time. *Manage Sci* **54**(1) 41–55.
- Paeratakul, S., D.P. Ferdinand, C.M. Champagne, D.H. Ryan, G.A. Bray. 2003. Fast-food consumption among US adults and children: dietary and nutrient intake profile. *Journal of the American Dietetic Association* **103**(10) 1332–1338.
- Petrin, A. 2002. Quantifying the benefits of new products: The case of the minivan. *Journal of Political Economy* **110**(4) 705–729.
- Png, IPL, D. Reitman. 1994. Service time competition. *The Rand Journal of Economics* 619–634.
- So, K. 2000. Price and time competition for service delivery. *Manufacturing Service and Operations Management* **2**(4) 392–409.
- Thomadsen, R. 2005a. The effect of ownership structure on prices in geographically differentiated industries. *The Rand Journal of Economics* **36**(4).
- Thomadsen, R. 2005b. Existence of pure strategy equilibria among geographically dispersed firms. *Economics Bulletin* **12**(12) 1–4.
- Topkis, D.M. 1998. *Supermodularity and complementarity*. Princeton University Press.
- Wang, H., T.L. Olsen. 2008. Service Capacity Competition with Peak Arrivals and Delay Sensitive Customers. *Working Paper. Washington University, St Louis, MO* .