Competition In Service Industries

Gad Allon
Graduate School Of Business
Columbia University
New York, NY, 10027

Awi Federgruen
Graduate School Of Business
Columbia University
New York, NY, 10027

October 2003, Revised: May 2004, Jan 2005; September 2005

Abstract

We analyze a general market for an industry of competing service facilities. Firms differentiate themselves by their price levels and the waiting time their customers experience, as well as different attributes not determined directly through competition. Our model therefore assumes that the expected demand experienced by a given firm may depend on *all* of the industry's price levels as well as a (steady state) waiting time standard, which each of the firms announces and commits itself to by proper adjustment of its capacity level. We focus primarily on a separable specification, which, in addition is linear in the prices. (Alternative non-separable or non-linear specifications are discussed in the concluding section.) We define a firm's service level as the difference between an upper bound benchmark for the waiting time standard ($\overline{w}$) and the firm's actual waiting time standard.

Different types of competition and resulting equilibrium behavior may arise, depending on the industry dynamics through which the firms select their strategic choices. In one case, firms may *initially* select their waiting time standards, followed by a selection of their prices in a *second* stage (Service Level First). Alternatively, the sequence of strategic choices may be reversed (Price First) or as a third alternative, the firms may make their choices simultaneously (Simultaneous Competition). We model each of the service facilities as a single server M/M/1 queueing facility, which receives a given firm specific price for each customer served. Each firm incurs a given cost per customer served as well as cost per unit of time proportional to its adopted capacity level.

Subject Classification
Queues, Multichannel
Games, Non-cooperative
Marketing, Competitive Strategy

1 Introduction and Summary

We analyze a general market for an industry of competing service facilities. Firms differentiate themselves by their price levels and the waiting time their customers experience, as well as different attributes not determined directly through competition. A given firm's demand volume may depend on all prices and all (steady state) waiting time standards in the industry. The latter may be specified as the *expected* steady state waiting time or a given (e.g. 95th) percentile of the waiting time distribution. In some settings, the waiting time standard is explicitly announced, possibly with monetary compensation offered if a customer's waiting time exceeds the standard. In other cases, it is the waiting time performance as observed by the clientele or reported by independent organizations. Either way, each firm commits itself to the chosen standard by adopting a sufficiently large capacity level. Different types of competition and equilibrium behavior arise, depending on the industry dynamics through which the firms make their strategic choices. In one case, the firms make all choices simultaneously: *Simultaneous Competition* (SC). Alternatively, firms may *initially* choose their waiting time standards, selecting their prices in a *second* stage: *Service Level First Competition* (SF). As a third alternative, the sequence of strategic choices may be reversed: *Price First Competition* (PF).

Numerous service industries use waiting time standards as an explicitly advertised competitive instrument. Dominos has offered free of charge delivery if pizza delivery were to take more than 30 minutes. Restaurant chains such as Black Angus offer free lunches if lunch is not served within 10 minutes. Banks like Wells Fargo award $5 when a customer waits more than five minutes in line. Various call or contact centers promise that the customer will be helped within one hour, possibly by a call back. Supermarket chains like Lucky launched a "3 is a crowd" campaign, guaranteeing that no checkout counter line would have more than three customers waiting. Ameritrade made major inroads into the online discounted brokerage market, waiving commissions for certain types of trades if service were to take more than 10 seconds. As a final example, airlines advertise waiting time characteristics such as "on time arrival percentage", while independent government agencies (e.g. the Aviation Consumer Protection Division of the DOT), 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.[1] Mazzeo (2001) shows that “on time arrival percentages” increase significantly with the number of competing carriers on the flight link.

Customers select a specific firm by trading off *three* categories of service attributes: (1) the *price* (2) the *waiting time standard* (3) all *other attributes.* For example, for overnight mail services, the ”other attributes” include the convenience of the pick-up process, the ease at which deliveries can be traced and the likelihood of the packages being damaged. In the restaurant and fast food industry, the location, ambiance and the quality of the food are important components of “other attributes”, and for internet service providers, the frequency of service interruption and the quality of the staff. Prior service competition models assume that the first two attributes (i.e. price and waiting time) can be aggregated into a so called *full price*, usually defined as the direct price plus a multiple of the expected waiting time. This is tantamount to assuming that all customers assign a *specific* cost value to their waiting time *and* that the cost of waiting is simply proportional to the total waiting time.

Many studies in the *modern* psychology, economics, marketing and operations literature have demonstrated that both assumptions are often violated. Kahneman and Tversky (1984)’s “calculator and jacket” experiment showed that the amount of time an individual is willing to spend so as to reduce an item’s purchase price by $1, varies drastically with the item’s base price. This experiment, confirmed by many other papers; (e.g., Leclerc et al (1995) and the references therein) shows that even on an *individual* level, no uniform waiting cost rate prevails. Carmon et al (1994) focus on the need to represent the cost of waiting as a non-linear function of the waiting time. Finally, Larson’s (1987) experiments show that the “disutility” of waiting varies in a highly non-linear manner with the customer’s delay as well as with many characteristics of the “queueing environment”.

The full price assumptions reduce the customers’ choice to a tradeoff between the full price and the ”other attributes”. Many prior models also assume that *all* customers select a firm with the *lowest* full price, albeit that different customers may be attracted

[1]Most industry observers recognize this as the dominant dimension of service. See e.g. Bowen and Headley (2001)

to different firms because of differences in their waiting time cost rate. This, of course amounts to assuming that the firms' services are perfect substitutes, i.e., *no* attributes, other than price and waiting time, matter, reducing the customers' multidimensional tradeoff process to the *full price* as the *single* criterion.

Defining a firm's *service level* as the difference between a given upper bound benchmark for the waiting time standard and the *actual* waiting time standard, we represent a firm's demand rate as a function of *all* prices and service levels in the industry. (We focus, primarily on a separable specification which, in addition, is linear in the price vector.) This class of demand models represents *general* tradeoffs between the above three categories of attributes. Price and waiting time are treated as truly independent attributes, in that, in general, a change in a firm's waiting time (distribution) can *not* be compensated for by a price change that leaves all firms' demand volumes unchanged. We model each firm as an $(M/M/1)$ queueing facility, which receives a given firm-specific price and incurs a given cost per customer served. Each firm incurs a cost per unit of time proportional to its adopted capacity level, determined to satisfy the waiting time standard under the expected demand rate.

We characterize the equilibrium behavior in the above three possible ways in which prices and service levels may be selected, i.e., (SC), (PF) and (SF). We show that, in all three settings, an equilibrium pair of price and service level vectors exists, in full generality, provided the waiting time benchmark is not excessively large. When characterizing the equilibrium behavior in these markets, we assume that the set of firms is given; in other words, we do not consider the possibility of firms exiting or entering the industry. We also develop efficient procedures to compute the equilibria in the various competition models.

These existence results are in stark contrast to the known behavior in existing service competition models. For example, the seminal model, due to Luski (1976) and Levhari and Luski (1978) confines itself to 2 service providers, assumes all customers choose their provider strictly on the basis of the full price, i.e., the price plus the expected waiting time multiplied with a customer specific cost rate. Customers' cost rates are independent and identically distributed (i.i.d). With service rates exogenously given, the competition between the two firms is confined to their price choices only. Whether an equilibrium exists in this elementary model remained an open question until it was answered in the *affirmative* by Chen and Wan (2000), for the case in which the firms' service rates are identical, while under non-identical service rates an example is given where *no* (pure) Nash equilibrium exists. The same example shows that the equilibrium behavior is very *unstable* : as the total market size varies from 1.2 to 1.3, the industry moves from a unique equilibrium to no-equilibrium to an infinite number of equilibria.

Cachon and Harker (2002), again for the case of two service providers, allows each firm's demand rate to be specified as a function of both firms' full price values; in this model, customers do not necessarily patronize the lowest full price provider (i.e., other attributes matter.) When the demand rate functions are linear, the known equilibrium results merely exclude the existence of multiple equilibria, and this only when the demand rates are sufficiently large. See Allon and Federgruen (2004). When the demand rate functions are (truncated) Logit functions, the authors examine a specific *symmetric* numerical instance. Varying a single cost rate parameter, the industry moves from a situation with a unique equilibrium under which both firms share the market, to one without any equilibrium, and next to a situation with two equilibria, one with firm 1 and the other with firm 2 as the monopoly provider.

To further appreciate the existence results for an equilibrium in the three competition models, note that they apply to an *arbitrary* number of competing service providers.[2] Also, in the (SC) model, the non-cooperative game involves *essentially multi-dimensional* strategy spaces.[3] Finally, in the (PF) and (SF) models, the existence results pertain to *two- stage* games. In the process of analyzing these two stage games, we characterize the price (service level) equilibrium which arises under a given vector of service levels (prices) and show how the former vary as a function of the latter. These second stage "price-only" and "service only" competition models are of interest by themselves, in settings in which one of the two strategic variables is specified in a way different than through non-cooperative competition.

[2]See e.g. Vives (2001), page 15, "non existence of a Nash equilibrium in pure strategies is pervasive in oligopoly models"

[3]A firm's strategy space is *essentially* multi-dimensional, if each of the strategy variables (e.g. price and service level) impacts on *all* firms' profit functions and these strategy variables cannot be replaced by a single aggregate variable (e.g. the full price)

We cannot guarantee that the equilibrium is unique. In general, the existence of multiple equilibria is unsettling, as it is hard to predict which of the equilibria is adopted by an industry. We show, however, that, in our model, the set of equilibria always has a *component-wise largest* and a *component-wise smallest* pair of equilibrium vectors. In other words, there exists an equilibrium such that each firm's price as well as its service level is *higher*, and there exists an equilibrium such that these are *lower* than his price and service level under any other Nash equilibrium. Most importantly, the component-wise largest pair of price and service level vectors is preferred by all of the firms. Finally, the schemes used to compute an equilibrium can also be applied to verify numerically whether multiple equilibria exist. Evaluating thousands of instances across a broad spectrum of parameters, we have never encountered a case with multiple equilibria.

The set of equilibria is *identical* under the (SC) and the (PF) models. Moreover, each firm's equilibrium service level in *any* such equilibrium is *uniquely* determined as a function of *that* firm's characteristics only, and it is a dominant choice for this firm, i.e., with fixed prices, the equilibrium service level is the firm's optimal choice, regardless of what service levels are adopted by its competitors. In contrast, the equilibrium in the (SF) model differs from that in the other two competition models. Here, a firm's equilibrium service level *does* depend, in general, on the characteristics of the competitors. Assuming the (SF) model has a unique equilibrium, we derive a simple sufficient condition under which each firm adopts a *higher price* and a *higher service level* while enjoying a *higher* demand volume, compared to the other types of competition. In the presence of multiple equilibria, the same uniform ranking applies to the component-wise smallest equilibria. Thus, if firms choose and announce their service levels before choosing their price, this will result in higher, but more expensive service by all competitors. Since all firms' demand volumes increase as well, this type of competition appears to benefit the consumer. It also suggests that value is added to the consumer when government agencies, industry consortia or independent organizations periodically report on service levels.

We briefly review several seminal papers (beyond those by Levhari and Luski and Cachon and Harker (2002), see above). We refer to Allon and Federgruen (2004) for a systematic discussion of several variants and extensions thereof and to Hassin and Haviv (2003) for a general survey of queueing models with competition. Kalai et. al. (1992)

initiated a stream of papers in which service firms compete in terms of their *capacity choices* with *exogenously* given prices, in contrast to the Luski (1976) and Levhari and Luski (1978) models in which firms compete in terms of their prices, with *fixed* capacity levels. They model the service industry as an M/M/2 system with two competing servers, i.e. all customers are served on a FIFO basis from a single queue. (If a customer arrives when both servers are idle, he is randomly assigned to one of them). The authors show that asymmetric Nash equilibria of service rate pairs may arise, sometimes associated with infinite waiting times.

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.

All of the above papers assume that customers either have no choice in selecting their service provider, or make the selection on the basis of the *full price* only. So (2000) and Cachon and Harker (2002), are the first to consider *differentiated* services, i.e., to analyze a model in which *other* service attributes matter along with the full price. In contrast to the latter, So (2000) establishes the existence of a unique equilibrium with an arbitrary number of competing M/M/1 service firms, when the demand rate functions are specified as a special type of attraction model. See Bell et al (1975). Here, each firm is characterized by an attraction value specified as a function of the firm's price and waiting time standard. With a fixed total market size, each firm's market share is given by the ratio of *its* attraction value and the *sum* of the industry's attraction values. So (2000) specifies the logarithm of the firms' attraction values as a *common* positive linear combination of the logarithms of the prices and the waiting time standards, plus a firm dependent constant. As in Cachon and Harker (2002) all firms' profit functions can be expressed as a function of the vector of full prices only; in So (2000), it is the attraction value which represents the full price. Afeche and Mendelson (2004) address a single firm model in which customers aggregate price and waiting time via a full price measure, now specified as a function of the price and *two* characteristics of the waiting time distribution. Ours appears to be the first competition model to address differentiated services while treating the prices and waiting time standards as fully independent attributes. This allows for different customers to exercise different explicit or implicit tradeoffs.

The model and notation are introduced in §2. The three competition models are analyzed in §3-§5. §6 establishes the above comparisons of the equilibrium in the three competition models. §7 completes the paper with numerical investigations and generalizations.

2 The Model

Consider a service industry with N competing service firms, each acting as an M/M/1 facility.[4] Each firm i positions itself in the market by selecting a price p_i as well as a service level θ_i. The latter may be defined in terms of the *expected* (steady state) waiting time $w_i = \mathbb{E}(W_i)$, or in terms of a given, say, ϕ *fractile* of the waiting time *distribution*, $w_i(\phi)$, $0 < \phi < 1$. For a given service rate μ_i and demand volume λ_i, it is known that

$$w_i = \frac{1}{\mu_i - \lambda_i}, \quad , w_i(\phi) = \frac{\ln(\frac{1}{1-\phi})}{\mu_i - \lambda_i} \tag{1}$$

(Note that $\mathbb{P}(W_i \leq w_i^0) = 1 - e^{-(\mu_i - \lambda_i) w_i^0}$, from which the expression for $w_i(\phi)$ in (1) readily follows.) The service level θ_i is defined as the difference between a *benchmark* upper bound $\overline{w}$ or $\overline{w}(\phi)$, and the actual waiting time standard w_i or $w_i(\phi)$ respectively, i.e. $\theta_i = \overline{w} - w_i$, or $\theta_i = \overline{w}(\phi) - w_i(\phi)$. For example, no internet access provider would offer an expected waiting time for access above one minute (say). Similarly, no contact contact would offer a guaranteed call back time above 24 hours. Thus, in these two examples, $\overline{w} = 1$ minute and $\overline{w} = 24$ hours could be used as the upper bound benchmark.

Each firm i is able to select its capacity or service rate so as to guarantee any given waiting time standard between 0 and $\overline{w}$ (or 0 and $\overline{w}(\phi)$ when standards are specified in terms of the ϕ fractile of the waiting time distribution). Thus, $\theta_i \in [0, \overline{w}]$. Assuming $\lambda_i > 0$, the required value of μ_i is easily obtained from (1):

$$\mu_i = \lambda_i + \frac{1}{w_i}, \quad \text{or} \quad \mu_i = \lambda_i + \frac{\ln(\frac{1}{1-\phi})}{w_i(\phi)} \tag{2}$$

[4]As mentioned, we do not model the possibility of new firms entering or firms exiting the market to pursue different opportunities.

(When $\lambda_i = 0, \mu_i = 0$ as well.) The two terms in (2) represent the two components the required capacity consists of: the first ,*volume based capacity*, is the base capacity ensuring that the service process is stable; the second component enables the desired waiting time standard and is referred to as the *service based capacity.*

Each firm i incurs a given cost c_i per customer served and a cost γ_i per unit of capacity, per unit of time. If the waiting time standard is based on the ϕ fractile of the waiting time distribution and firm i offers to pay a penalty C_i to any customer whose waiting time is in excess of the stated $w_i(\phi)$, this adds an expected cost per customer $(1-\phi)C_i$. Such penalties are therefore easily incorporated into the analysis, simply by replacing c_i by $\widehat{c}_i = c_i + (1-\phi)C_i$.

The price p_i may be chosen from an interval $[p_i^{min}, p_i^{max}], i = 1, \ldots, N$. Clearly, firm i selects a price p_i which results in a non-negative gross profit margin $p_i - c_i - \gamma_i$. (By (2), $c_i + \gamma_i$ is the marginal cost per unit of demand.) Thus, without loss of generality, we select

$$p_i^{min} = c_i + \gamma_i, \quad i = 1, \ldots, N. \tag{3}$$

As to p_i^{max}, it is chosen to be sufficiently large as to have no impact on the equilibrium behavior. In full generality, the demand rates would be specified as general functions of *all* prices and waiting time standards (i.e. $\lambda_i = \lambda_i(p, \theta)$), that obey obvious monotonicity properties. We focus primarily on specifications in which the demand rate (,when positive,)is a separable function of the prices and service levels, which, in addition, is linear in the price vector:

$$\lambda_i(p, \theta) = \left[a_i(\theta_i) - b_i p_i - \sum_{j \neq i} \alpha_{ij}(\theta_j) + \sum_{j \neq i} \beta_{ij} p_j \right]^+ \tag{4}$$

where $x^+ = \max(x, 0)$. This quasi-separable specification enables tractable analyses and estimation procedures, as with standard linear equations. The functions $a_i(\theta_i)$ are assumed to be three times differentiable, increasing, and concave in the service level θ_i , i.e., equal size reductions in the waiting time standard result in progressively smaller increases of the demand volume. As to the cross term functions $\alpha_{ij}(\theta_j)$, they are merely assumed to be non-decreasing and differentiable. Without loss of practical generality, we assume that a *uniform* price increase by all N firms cannot result in an increase in any firm's demand volume, and that a price increase by a given firm can not result in an increase of the industry's aggregate demand volume, i.e.

$$(D) \quad b_i > \sum_{j \neq i} \beta_{ij}, i = 1, \ldots, N; \qquad (D') \quad b_i > \sum_{j \neq i} \beta_{ji}, i = 1, \ldots, N$$

This condition is usually referred to as the "Dominant Diagonal" condition.

Alternative specifications of the demand functions include

(i) the attraction models (ATT): $\lambda_i(p, \theta) = \frac{v_i(p_i, \theta_i)}{\sum_{j=1}^N v_j(p_j, \theta_j) + v_0}$, with v_0 a positive constant and $v_i(\cdot, \cdot)$ a function which is decreasing in its first argument and increasing in its second. Within the latter broad class of models, supported by axiomatic foundations, it is prevalent to choose a log-separable specification of the attraction value v_i, i.e. $v_i(p_i, \theta_i) = \phi_i(p_i)\psi_i(\theta_i)$, a natural extension of the Multinomial Logit (MNL) model;[5]

(ii) the natural extension of the Cobb - Douglas specification (CD): $\lambda_i(p, \theta) = \frac{a_i(\theta_i)}{\prod_{j \neq i} \alpha_{ij}(\theta_j)} p_i^{-b_i} \prod_{j \neq i} p_j^{\beta_{ij}}$

(iii) Generalized CES models: $\lambda_i = M \frac{p_i^{b_i - 1} \theta_i a}{\sum_{j=1}^N p_j^{b_j} \theta_j^{a+1}}, i = 1, \ldots, N$.[6] In §6, we discuss how our results carry over to these classes of non-linear demand models.

Under all of the above specifications (i)-(iii), a firm maintains a *positive* market share irrespective of how extreme and uncompetitive its price and service level choices are. In contrast, (4), in addition to enjoying analytical simplifications, specifies a firm's demand to be zero under such extreme choices; this appears more realistic in most industries. We show that in all of the competitive models considered, the firms' *equilibrium* choices induce a positive market share for each. To guarantee that this is the case , it suffices in the Price and Service Level First models to assume,

$$\lambda_i(c + \gamma, \theta) > 0, i = 1, \ldots, N, \forall \theta \in [0, \overline{w}]^N, \tag{5}$$

i.e., any firm i can achieve a positive market share, at least when willing to operate with a zero variable profit margin, i.e. when $p_i = p_i^{min} = c_i + \gamma_i$. (5) guarantees that under *this* price, $\lambda_i > 0$, regardless of the competitors' choices. [7] However, other price-service

[5] See Bell et. al. (1975)

[6] The experiments in Kahneman and Tversky (1984) show that the amount of time a typical consumer is willing to add to the waiting time by switching to an alternative provider depends primarily on the *relative* rather than the *absolute* price reduction the switch accomplishes. The (CD) and (CES) models are well suited to reflect this phenomenon

[7] (5) reduces to lower bounds for the intercept values $a_i(0) > \sum_{j \neq i} \alpha_{ij}(\overline{w}) + b_i(c_i + \gamma_i) - \sum_{j \neq i} \beta_{ij}(c_j + \gamma_j)$.

level combinations may result in zero demand. In the two remaining competition models (Simultaneous Competition and Price First competition), a somewhat stronger condition is needed, namely:

$$\lambda_i(p,\theta) > 0, \forall p \in \times_{i=1}^{N}\left[p_i^{min}, p_i^{max}\right], \theta \in [0,\overline{w}]^{N}\ ^{8} \tag{6}$$

(As mentioned, (6) is satisfied, without any parameter restrictions, for any of the non-linear demand functions mentioned above.)

As is well known from the literature on oligopoly models with *product differentiation*, systems of demand equations need not, but often can be obtained from one of several underlying consumer utility models, in particular the *representative consumer model*, the *random utility model* and the *address model*. See e.g. Anderson et. al. (1992). Similarly, (4) may, e.g. be derived from a representative consumer model with utility function $U(\lambda,\theta) \equiv C + \frac{1}{2}\lambda^T B^{-1}\lambda - \lambda^T B^{-1}\overline{a}(\theta)$ where the $N \times N$ matrix B has $B_{ii} = -b_i$ and $B_{ij} = \beta_{ij}, i \neq j$, $\overline{a}(\theta) \equiv a_i(\theta_i) - \sum_{j\neq i}\alpha_{ij}(\theta_j)$ and $C > 0$. ((D) ensures that B^{-1} exists and is negative semi-definite, giving rise to a jointly concave utility function). The demand functions (4) arise by optimizing the utility function subject to a budget constraint.

Thus, if the waiting time standards are expressed in terms of the *expected* waiting time, each firm i's long run average profit Π_i for $i = 1,\ldots,N$ is given by the function:

$$\Pi_i(p,\theta) = \begin{cases} \lambda_i(p_i - c_i - \gamma_i) - \frac{\gamma_i}{\overline{w}-\theta_i} & \text{,if } \lambda_i > 0 \\ 0 & \text{,otherwise} \end{cases} \tag{7}$$

A firm i may thus avoid a loss by adopting a sales volume $\lambda_i = 0$. Losses may, in principle, still occur when $\lambda_i > 0$, in case the cost associated with the service based capacity dominates the gross profits. As mentioned in Footnote 4, we do not consider the possibility of firms existing the industry; in particular, we do not impose a participation constraint for the firms.

If the waiting time standard is expressed in terms of the ϕ fractile of the waiting time distribution, the profit functions Π_i are identical to those in (7) except that the last term to the right of (7) is given by $\frac{\gamma_i \ln(\frac{1}{1-\phi})}{\overline{w}(\phi)-\theta_i}$. In view of the close similarity between the profit functions under the expected waiting time and waiting time fractile-based standards, we henceforth confine ourselves to the former case. Finally, one may envision settings in

[8] (6) reduces to similar lower bounds for the intercept values $a_i(0)$.

which customers are sensitive to *both* the expected waiting time *and* a given ϕ fractile of the waiting time distribution, giving rise to demand equations of the form

$$\lambda_i = \left[a_i^E(w_i) + a_i^T(w_i(\phi)) - b_i p_i - \sum_{j \neq i} \left(\alpha_{ij}^E(w_j) + \alpha_{ij}^T(w_j(\phi)) \right) + \sum_{j \neq i} \beta_{ij} p_j \right]^+ . \quad (8)$$

Since $w_i(\phi) = w_i \ln\left(\frac{1}{1-\phi}\right)$, (8) is equivalent to (4) with $a_i(\theta_i) \equiv a_i^E(\overline{w} - \theta_i) + a_i^T\left(\ln\left(\frac{1}{1-\phi}\right)(\overline{w} - \theta_i)\right)$ and $\alpha_{ij}(\theta_j) \equiv \alpha_{ij}^E(\overline{w} - \theta_j) + \alpha_{ij}^T\left(\ln\left(\frac{1}{1-\phi}\right)(\overline{w} - \theta_j)\right)$. This general setting with customers sensitive to *multiple* waiting time standards can thus be reduced, without loss of generality, to a model with a *single* waiting time standard.

As mentioned in §1, Cachon and Harker (2002) consider, in the case of a *duopoly*, a demand model of the form $\lambda_i = A_i - b_i F_i + \beta_{ij} F_j$, $j \neq i$, where F_i = "the full price" paid by customers of firm i, i.e. $F_i = p_i + k w_i$. [9] This specification is based on *two* important assumptions: *first*, all potential customers aggregate the price and waiting time standards into a *single* aggregate measure ("the full price"). *Second*, every unit of time waited has the same dollar value k for all potential customers, regardless of how long the total waiting time is. We retrieve the full price model from the general demand model (4) by adopting the following special choices:

$$a_i(\theta_i) = a_i^0 + a_i^1 \theta_i; \alpha_{ij}(\theta_j) = \alpha_{ij}^1 \theta_j; i \neq j \quad (9)$$

$$a_i^1 = k b_i \quad ; \alpha_{ij}^1 = k \beta_{ij}; \quad i \neq j. \quad (10)$$

In other words, all intercept functions are affine and their slopes proportional to the price effects b and β with k as the "common proportionality factor". [10] Unless the special relationships in (9) and (10) apply, prices and waiting time standards function as truly independent attributes. For example, under affine $a_i(\cdot)$ and $\alpha_{ij}(\cdot)$ functions, as in (9), if firm i decreases its service level by 1 unit, it must decrease its price by $\frac{a_i^1}{b_i}$ units to leave its demand volume unchanged. The net effect on firm j's demand volume is then an increase by $\beta_{ji} \frac{a_i^1}{b_i} - \alpha_{ji}^1 \neq 0$, unless (10) applies.

[9] This paper considers also non-linear functions of F_1 and F_2

[10] The "winner take all" setting in which the firm with the lowest full price captures the complete market, does not arise as a special case of the general model (4), even though the parameters in (4) can be chosen to reflect settings in which small differences in the full price can result in very large differences in market share.

Since the full price model arises as a special case of the general model, all of our characterizations of the equilibrium behavior which apply to the latter, apply, a fortiori, to the former. This resolves an outstanding question in Cachon and Harker (2002), whether an equilibrium exists, even in the case of a *duopoly*. Our results for the attraction models (ATT) confirm the possibility of non-existence of equilibria, as pointed out by Cachon and Harker (2002), while establishing conditions under which the existence of an equilibrium is guaranteed.

If the full price F_i is treated as the single strategic instrument of firm i, this precludes the modeling of settings where prices and waiting times are selected sequentially. In §1, we discussed various a priori reasons and experimental results which challenge the applicability of the full price model in specific settings. In addition an upfront restriction to the full price model also invokes difficulties when estimating the system of demand functions. First, when specifying a *linear* dependence on the prices, it is, in the full price model, necessary to restrict oneself to affine $a(\cdot)$ and $\alpha(\cdot)$ - functions as well. Moreover, while the firms' prices and expected waiting times can be observed directly and used as explanatory variables in a system of regression equations, the full price values are not observable, as they depend on the value of k. It is, of course, possible to estimate the parameters in (4) under the constraints imposed by (9) and (10), but these constraints add significant (and apparently unnecessary) difficulties to the estimation procedure. Finally, numerical experiments in Allon and Federgruen (2004) suggest that the imposition of the parameters constraints (10) may result in very significant changes in the associated equilibria. In particular, Allon and Federgruen (2004) show how starting with a model which satisfies (9) and (10), i.e. which belongs to the full price model, large deviations in, for example, the SC equilibria can be observed, resulting from relatively small deviations in the ratios $\left\{\frac{a_i'}{b_i}; \frac{\alpha_{ij}'}{\beta_{ij}} : i \neq j\right\}$ from their common value, or when appending non-linear terms to the $a(\cdot)$ and $\alpha(\cdot)$ functions.

3 Service Level First Model

Often, firms face significantly higher stickiness for their service level choices as compared to their ability to vary prices, or vice versa. Relative stickiness of the former may, for example, arise because of human resource practices, labor contracts or long lead times for technology purchases. Recall that a firm's required capacity consists of *two* components; since the service based component only depends of the firm's own service level, at least this part of the capacity investment can be fixed over a larger horizon by selecting and maintaining a given service level. In the airline industry, we observe that reservation call centers are typically designed to handle 80% of the Economy Class passengers within 20 seconds. Airlines have stuck with this waiting time standard for years while willing to change prices daily. Conversely, some industries experience a higher level of price rigidity. See Blinder et al (1998), and the many references therein, for a comprehensive theoretical and empirical investigation of 12 factors underlying price stickiness.

To address industries with relatively higher service level rigidity we analyze, in this section, the (SF) model. In the next section, we analyze the (SC) model where firms select or adapt their prices and service levels simultaneously. This assumption is valid when it is equally easy or difficult to adapt either one of the two strategic dimensions. Finally, §5 analyzes the (PF) model to characterize settings with higher *price* rigidity. To analyze the two stage game (SF), we start with the second stage price game which arises under a given vector of service levels θ^0.

3.1 Price Competition Model

We show that the price-competition game has a *unique* price equilibrium p^*, which satisfies the first order conditions

$$\frac{\partial \Pi_i}{\partial p_i} = -b_i(p_i - c_i - \gamma_i) + \lambda_i. \tag{11}$$

In matrix notation, this linear system of equations can be written in the form

$$Ap = \overline{a}(\theta) + \kappa \tag{12}$$

where the $N \times N$ matrix A is specified by $A_{ii} = 2b_i$, $A_{ij} = -\beta_{ij}, i \neq j$ and where $\kappa_i = b_i(c_i + \gamma_i)$. We first state the following properties of the matrix A which were shown in parts (a)-(c) of Lemma 2 in Bernstein and Federgruen (2002):

Lemma 1. (a) *A is invertible, $A^{-1} \geq 0$ and every entry of A^{-1} is nondecreasing in each of the β_{ij} coefficients.*

(b) *Let* $\delta_i \equiv b_i(A^{-1})_{ii}$*. Then,* $0.5 \le \delta_i < 1$

(c) $(A^{-1})_{ij} \le \frac{1}{b_j}$

We refer to δ_i as the *degree of positive externality* faced by firm i and note from Lemma 1 that it is a dimensionless index which varies between 0.5 and 1 and increases with each of the β coefficients. The following theorem characterizes the equilibrium in the price competition model and shows how the equilibrium prices and demand rates respond to changes in the cost parameters and service levels. The proof uses the theory of supermodular games. A function $f : \mathbb{R}^N \to \mathbb{R}$ is supermodular is it has the increasing difference property:

$$f(x_i^1, x_{-i}) - f(x_i^2, x_{-i}), \text{ increases in } x_{-i} \text{ for all } x_i^1 > x_i^2 \tag{13}$$

Theorem 1. *(Price Competition Model) Fix a service level vector* θ *and assume condition (5) applies.*

(a) *The price competition game has a unique equilibrium which satisfies (12), i.e.* $p^*(\theta) = A^{-1}(\overline{a}(\theta) + \kappa) \in (p^{min}, p^u \equiv A^{-1}(a(\overline{w}) + \kappa))$*. The equilibrium demand volume for firm* i *is given by* $\lambda_i^* \equiv \lambda_i(p^*) = b_i(p_i^* - c_i - \gamma_i) > 0$ *and the equilibrium profit for firm* i *is given by* $\pi_i^* = b_i\left(p_i^* - c_i - \gamma_i\right)^2 - \frac{\gamma_i}{\overline{w} - \theta_i}$

(b) p^* *and* λ^* *are increasing in each of the cost parameters* $\{c_i, \gamma_i, i = 1, \ldots, N\}$ *with* $\frac{\partial p_i^*}{\partial c_i} = \frac{\partial p_i^*}{\partial \gamma_i} = \delta_i$ *and* $\frac{\partial \lambda_i^*}{\partial c_i} = \frac{\partial \lambda_i^*}{\partial \gamma_i} = b_i(\delta_i - 1) < 0, \quad i = 1, \ldots, N$

(c)

$$\frac{\partial p_i^*}{\partial \theta_j} = \frac{1}{b_i}\frac{\partial \lambda_i^*}{\partial \theta_j} = (A^{-1})_{ij} a_j'(\theta_j) - \sum_{l \ne j} (A^{-1})_{il} a_{lj}'(\theta_j). \tag{14}$$

When the cross term functions $\alpha_{ij}(\cdot)$ *are linear or convex, each equilibrium price and volume is a separable concave function of* θ*.*

Proof. **(a)** Under a given service level vector θ, each firm i has committed to a given positive service level based capacity $\frac{1}{w_i} = \frac{1}{\overline{w} - \theta_i}$. Consider the modified game $\overline{\mathcal{G}}$ with profit functions

$$\overline{\Pi}_i = \left(a_i(\theta_i) - b_i p_i - \sum_{j \ne i} \alpha_{ij}(\theta_j) + \sum_{j \ne i} \beta_{ij} p_j \right) (p_i - c_i - \gamma_i) - \frac{\gamma_i}{\overline{w} - \theta_i}. \tag{15}$$

We first show that $\overline{\mathcal{G}}$ has a unique equilibrium $p^*(\theta)$ with $\lambda_i > 0, i = 1, \ldots, N$. $\frac{\partial^2 \overline{\Pi}_i}{\partial p_i \partial p_j} = \beta_{ij} \geq 0$, i.e., the profit function $\overline{\Pi}_i$ is supermodular, in (p_i, p_j), see (13). Since the feasible action set of each firm is a closed interval, the game is supermodular and possesses an equilibrium. The fact that it has a *unique* equilibrium follows from (D). See e.g. Milgrom and Roberts (1990). Note that each profit function $\overline{\Pi}_i$ is concave in the price variable p_i; thus, if the first order conditions (12) have a solution in the feasible set $[p^{min}, p^{max}]$, this solution must be the unique equilibrium. But (12) has the solution $p^{min} = c + \gamma \leq p^*(\theta) = A^{-1}(a(\theta) + \kappa) \leq A^{-1}(a(\overline{w}) + \kappa)$. (To verify the first inequality, (5) implies $\overline{a}(\theta)_i + b_i(c_i + \gamma_i) + \sum_{j \neq i} \beta_{ij}(c_j + \gamma_j) > 2b_i(c_i + \gamma_i) \Rightarrow \overline{a}(\theta)_i + \kappa_i + > 2b_i(c_i + \gamma_i) - \sum_{j \neq i} \beta_{ij}(c_j + \gamma_j), i = 1, \ldots, N$. Thus, in matrix notation, $\overline{a}(\theta) + \kappa > A(c + \gamma) \Rightarrow A^{-1}(\overline{a}(\theta) + \kappa) > c + \gamma$, by Lemma 1(a). The second inequality is immediate from the properties of the $a-$ and $\alpha-$ functions.) Rewriting (12) and using (4) we obtain $\lambda_i^* = b_i(p_i^* - c_i - \gamma_i) > 0$.

To show that p^* is an equilibrium in the original game $\mathcal{G}$, note that $\Pi_i(p_i, p_{-i}^*, \theta) \leq \Pi_i(p^*, \theta)$ for all $p_i \in [p_i^{min}, p_i^{max}]$. If $\lambda_i(p_i, p_{-i}^*, \theta) = 0, \Pi_i(p_i, p_{-i}^*, \theta) = \frac{-\gamma_i}{\overline{w} - \theta_i} \leq \overline{\Pi}_i(p^*)$ where the inequality follows from $\lambda_i^* > 0$ and the equality from the fact that firm i pre-committed to the service based capacity $\frac{1}{\overline{w} - \theta_i}$. If $\lambda_i(p_i, p_{-i}^*, \theta) > 0, \Pi_i(p_i, p_{-i}^*, \theta) = \overline{\Pi}_i(p_i, p_{-i}^*, \theta) \leq \overline{\Pi}_i(p_i^*, \theta) = \Pi_i(p_i^*, \theta)$.

It remains to be shown that $\mathcal{G}$ has no other equilibrium. But, any equilibrium $\widetilde{p}$ with $\lambda_i(\widetilde{p}, \theta) > 0$ for $i = 1 \ldots, N$ must satisfy the first order conditions (12), with $p^*(\theta)$ as its unique solution. Thus $\widetilde{p} \neq p^*(\theta)$ must have $\lambda_i = 0$ for some $i = 1 \ldots, N$. Since $\lambda_i([c_i + \gamma_i, \widetilde{p}_{-i}], \theta) > 0$ and since λ_i is a continuous function of the price vector, there exists a price $\widetilde{p}_i \neq \widehat{p} > c_i + \gamma_i$ such that $\lambda_i([\widehat{p}, \widetilde{p}_{-i}], \theta) > 0$. Thus, firm i can increase the first term in the profit equation (7), without changing the second term, by switching to the price $\widehat{p}$. (The second term does not change because the firm is pre-committed to the service based capacity $\frac{1}{\overline{w} - \theta_i}$.) This contradicts the assumption that $\widetilde{p}$ is an equilibrium.

Substituting $\lambda_i^* = b_i^*(p_i^* - c_i - \gamma_i)$ into (7), we obtain:

$$\pi_i^*(\theta) = \Pi_i(p^*(\theta), \theta) = b_i(p_i^* - c_i - \gamma_i)^2 - \frac{\gamma}{\overline{w} - \theta_i}, i = 1, \ldots, N \tag{16}$$

(b) The fact that each equilibrium price p_i^* is increasing in each of the cost parameters $\{c_j, \gamma_j, j = 1, \ldots, N\}$ is immediate from $p^* = A^{-1}(\overline{a}(\theta) + \kappa)$ since $A^{-1} \geq 0$. Moreover, $\frac{\partial p_i}{\partial c_i} = \frac{\partial p_i}{\partial \gamma_i} = b_i(A^{-1})_{ii} = \delta_i$ by Lemma 1(b). Finally, it follows from part (a) that $\frac{\partial \lambda_i}{\partial c_i} = \frac{\partial \lambda_i}{\partial \gamma_i} = b_i\left(\frac{\partial p_i}{\partial \gamma_i} - 1\right) = b_i\left(\frac{\partial p_i}{\partial c_i} - 1\right) = b_i(\delta_i - 1)$.

(c) Immediate from part (a).

□

We conclude that p^u is a uniform upper bound for all feasible price equilibria. We henceforth assume

$$p^{max} \geq p^u, \tag{17}$$

specifying *explicitly* which choices of p^{max} are sufficiently large so as not to influence the equilibrium behavior.

Thus, if one of the cost parameters c_i or γ_i of firm i increases, the equilibrium price p_i^* increases by at least half as much, but never more than by the increase in the cost parameter itself. Moreover, the marginal price increase is given by the firm's degree of positive externality δ_i and is therefore increasing in any of the β-coefficients. For a fixed vector of direct price effects b, we observe from part (b) of Theorem 1 that under larger β coefficients, hence under a larger value for firm i's degree of positive externality δ_i, this firm is willing to make a bolder price adjustment to any increase in its cost parameters, thereby maintaining a larger portion of its original profit margin. The reason is that the firm's competitors respond with larger price increases themselves. Part (c) implies the existence of a critical value $0 \leq \theta_{ij}^0 \leq \overline{w}$ such that as firm j increases its service level, p_i^* and λ_i^* are increasing on the interval $[0, \theta_{ij}^0)$, and decreasing on $[\theta_{ij}^0, \overline{w})$.

Below we provide a simple and broadly satisfied condition under which the equilibrium price p_i^* and equilibrium demand volume λ_i^* vary *monotonically* with any of the service levels, i.e. $\theta_{ij}^0 = \overline{w}$ for all i, j.

Just like the equilibrium prices and volumes vary unimodally, and often monotonically with any of the service levels, the same can be said about the dependence of the firm's equilibrium profits on any of its competitors' service levels: it follows from (16) that

$$\frac{\partial \pi_i^*}{\partial \theta_j} = 2b_i(p_i^* - c_i - \gamma_i)\frac{\partial p_i^*}{\partial \theta_j}. \tag{18}$$

In particular, firm i's profit increases as a result of the service level improvement by a competing firm j if and only if the service level improvement results in an increase in firm i's price. Thus π_i^* increases on the interval $[0, \theta_{ij}^0]$ and decreases on the remaining interval $(\theta_{ij}^0, \overline{w}]$. As far as the dependence of firm i's equilibrium profit π_i^* on its own service level is concerned, a less clear cut picture emerges. The following general statement can however be made: since $\frac{\partial \pi_i^*}{\partial \theta_i} = 2b_i(p_i^* - c_i - \gamma_i)\frac{\partial p_i^*}{\partial \theta_i} - \frac{\gamma_i}{(\overline{w}-\theta_i)^2}$, two cases prevail. If increasing the firm's service level from $\theta_i = 0$ to any positive service level results in a price decrease, the firm's equilibrium profits are a decreasing function of its service level throughout, i.e. the firm is best off providing minimal service (independent of the service level choices of any of its competitors.) The second case arises when an increase from $\theta_i = 0$ to a marginally positive level allows the firm to charge a marginally *higher* price. In this case, a value $\theta_i^{pr} < \theta_{ii}^0$ exists, such that profits decrease when the service level exceeds θ_i^{pr}. On the other hand, on the interval $[0, \theta_i^{pr})$, the equilibrium profit may, in general, alternate arbitrarily between being increasing and decreasing. However, if all intercept functions $\{a_i(\cdot)\}$ and all cross term functions $\{\alpha_{ij}\}$ are affine, $\frac{\partial \pi_i^*}{\partial \theta_i}$ can be shown to be concave so that π_i, viewed as s function of θ_i, possesses at most two local optima in the interval $[0, \overline{w}]$. Combined with the fact that $\lim_{\theta_i \uparrow \overline{w}} \frac{\partial \pi_i^*}{\partial \theta_i} < 0$, this reveals that only one of three patterns may emerge: (i) profits decline throughout the feasible service level interval $[0, \overline{w}]$; (ii) profits are unimodal and (iii) profits first decline, then increase, and after reaching a local maximum, proceed to decline.

3.2 The Service Level First Model: The Two Stage Game

We now turn to the first stage game in which firms first select their service levels.

Theorem 2. *(The two stage game) Assume (5) holds. Assume $\overline{w}$ is sufficiently small* [11] *and the cross term functions $\{\alpha_{ij}, j \neq i\}$ are linear or convex. The Service Level First model has an equilibrium θ^* (and associated price equilibrium $p^*(\theta^*)$.)*

Proof. Fix $i = 1, \ldots, N$. Note from Theorem 1(c) that $\frac{\partial p_i^*}{\partial \theta_i}$ is independent of the service

[11] An upper bound for $\overline{w}$ can be derived as the root of a non-linear equation, see (19). When the $a(\cdot)$ and $\alpha(\cdot)$ functions are linear, $\overline{w} \leq \sqrt[3]{\frac{\gamma_i}{b_i\left[A_{ii}^{-1}a_i' - \sum_{l\neq i} A_{il}^{-1}\alpha_{il}'\right]^{-1}}}$

level choices of firm i's competitors; thus $\frac{\partial p_i^*}{\partial \theta_i} = p_i^{*'}(\theta_i)$. By (14),

$$\frac{\partial \pi_i^*}{\partial \theta_i} = 2b_i(p_i^* - c_i - \gamma_i)p_i^{*'}(\theta_i) - \frac{\gamma_i}{(\overline{w} - \theta_i)^2} \tag{19}$$

and

$$\frac{\partial^2 \pi_i^*}{\partial \theta_i^2} = 2b_i(p_i^* - c_i - \gamma_i)p_i^{*''}(\theta_i) + 2b_i\left(p_i^{*'}(\theta_i)\right)^2 - \frac{2\gamma_i}{(\overline{w} - \theta_i)^3}. \tag{20}$$

The first term in (20) is negative under linear or convex cross term functions $\{\alpha_{ij}, j \neq i\}$. Thus π_i^* is concave in θ_i, provided that $2b_i(p_i^{*'}(\theta_i))^2 \leq \frac{2\gamma_i}{(\overline{w}-\theta_i)^3}$, i.e. provided,

$$2b_i(p_i^{*'}(\theta_i))^2 \leq \frac{2\gamma_i}{\overline{w}^3} \Longleftrightarrow \overline{w} \leq \min_{\theta_i} \sqrt[3]{\frac{\gamma_i}{b_i(p_i^{*'}(\theta_i))^2}} = \min\left\{\sqrt[3]{\frac{\gamma_i}{b_i(p_i^{*'}(0))^2}}, \sqrt[3]{\frac{\gamma_i}{b_i(p_i^{*'}(\overline{w}))^2}}\right\} \tag{21}$$

where the last equality follows from $p_i^{*'}(\theta_i)$ being decreasing. Finally, the fact that $p_i^{*'}(\theta_i)$ is decreasing shows that condition (21) is satisfied for $\overline{w}$ sufficiently small. □

Arbitrary small or large utilizations rates may arise, for any value of $\overline{w}$ (firm i's utilization rate is given by $\frac{\lambda_i}{\lambda_i + w_i^{-1}} \leq \frac{\lambda_i}{\lambda_i + \overline{w}_i^{-1}}$.) The upper bound for $\overline{w}$ is required to guarantee that each of the first-stage profit functions π_i^*, is concave. Our numerical investigations show, however, that the qualitative properties of equilibrium are maintained, even if $\overline{w}$ is chosen at an arbitrarily large value. Similar observations apply to the competition models in §4-5, where a similar upper bound for $\overline{w}$ is required.

Theorem 2 does not guarantee that the two-stage game has a *unique* equilibrium. In the following Corollary, we (a) characterize the set of Nash equilibria, (b) show that it has a component-wise largest and a component-wise smallest element $\overline{\theta}_{SF}, \underline{\theta}_{SF}$ respectively, and (c) show that the following simple tatônnement scheme converges to $\overline{\theta}_{SF}(\underline{\theta}_{SF})$ when started at $0([\overline{w}, \ldots, \overline{w}]^T)$: starting with an arbitrary service level vector θ^0, determine in the k-th iteration of the scheme, θ^k such that $\forall i = 1, \ldots, N$ $\theta_i^k = \arg\max_{\theta_i} \pi_i^*(\theta_i, \theta_{-i}^{k-1})$, i.e. θ_i^k represents firm i's best response if all competitors adopt service levels from the vector θ^{k-1}. All three results are obtained by demonstrating, under condition (22) below, (but without any condition on $\overline{w}$,) that the first stage game is *supermodular*. Since the feasible ranges are closed intervals, the game is supermodular if and only if each of the reduced profit functions $\pi_i^*(\theta)$ is *supermodular*.

It follows from (18) that π_i^* is twice differentiable, so that (13) is satisfied if and only if $\frac{\partial^2 \pi_i^*}{\partial \theta_i \partial \theta_j} = 2b_i \frac{\partial p_i^*}{\partial \theta_i} \frac{\partial p_i^*}{\partial \theta_j} \geq 0$. The first stage game is thus supermodular if $\frac{\partial p_i^*}{\partial \theta_j}$ has a *uniform* sign for all i, j, throughout the feasible service level region $\times_{i=1}^{N}[0, \overline{w}]$. As discussed above, in general, a firm's equilibrium price fails to be monotone either in its own service level or that of any of its competitors. By (14), $\frac{\partial \lambda_i^*}{\partial \theta_j} \geq 0 \Leftrightarrow \frac{\partial p_i^*}{\partial \theta_j} \geq 0$ which itself is equivalent to

$$(A^{-1})_{ij} a_j'(\theta_j) \geq \sum_{l \neq j} (A^{-1})_{il} \alpha_{lj}'(\theta_j), \quad \forall i, j = 1, \ldots, N. \tag{22}$$

In other words, condition (22) requires that the *direct* impact of a service level improvement by firm j on its own demand volume be as large as a linear combination of the indirect effects this service level improvement has on the demand volumes of the other firms. This condition bears resemblance to (D') which states that the direct impact of a *price* increase by a firm on its own demand volume is at least as large as the sum of the indirect effects the price increase has on the demand volumes of the competitors. (Recall that (D') is equivalent to the highly plausible assumption that a price increase by one of the firms cannot result in an increase of the aggregate sales in the industry.) Moreover,to the extent that the indirect service sensitivities $\{\alpha_{kj}', k \neq j\}$ *are significant* compared to the direct service level sensitivity a_j', this reflects a highly competitive industry and is likely to be accompanied with indirect price sensitivities $\{\beta_{kj}\}$ being relatively large, compared to the direct price sensitivity b_j. However , the inequalities $(A^{-1})_{ij} a_j'(\theta_j) \geq \sum_{k \neq j} \frac{\alpha_{kj}'(\theta_j)}{b_k}$ are sufficient for (22). These inequalities are all the more easily satisfied as (any of) the β coefficient(s) increase(s) since each of the entries of the matrix A^{-1} is increasing in each of the β coefficients, see Lemma 1(a).

Since the first stage game is supermodular under (22), the following Corollary follows from Topkis (1998)

Corollary 1. *(Service Level First model: characterizations and computation of equilibria) Assume (5) and (22) holds.*

(a) *The first stage game in the Service Level First model is supermodular and it has a component-wise smallest and component-wise largest equilibrium $\underline{\theta}_{SF}, \overline{\theta}_{SF}$ (with associated price equilibria $p^*(\underline{\theta}_{SF}), p^*(\overline{\theta}_{SF})$).*

(b) *When starting at $\theta^0 = 0$ ($[\overline{w}, \overline{w}, \ldots, \overline{w}]^T$) the tatônnement scheme generates an increasing (decreasing) sequence of service level vectors which converges to $\underline{\theta}_{SF}(\overline{\theta}_{SF})$.*

In the $k+1$st iteration of the scheme, each firm i determines the value of θ which maximizes $\pi_i^*(\theta_i|\theta_{-i}^k) = b_i(p_i^*(\theta_i, \theta_{-i}^k) - c_i - \gamma_i)^2 - \frac{\gamma_i}{\overline{w}-\theta_i}$.

In the two stage competition model (SF), firms choose their prices after all service levels are revealed. In this setting, condition (22) has far reaching implications: not only does it guarantee that a service level improvement by firm j results in price *increases* by *all* of the firms, but the same applies to their volumes, see the first equality in (14).Even the demand volume of a competing firm i increases, due to the fact that the positive impact of firm i's price increase (along with those of the other firms $j \neq i$) on λ_i^* dominates the negative impact resulting from the increase in the cross term $\alpha_{ij}(\theta_j)$ and that of p_i^*, see (4). In addition, a service level improvement by a firm results in a profit improvement for all of its competitors, see (18), since it positively impacts on *both* their price *and* their demand volume, without changing their cost structure. At the same time, the impact of a service level improvement on the firm's own profits remains ambiguous, as in the case of the general model. However, for affine functions $\{a_i\}, \{\alpha_{ij}\}$, since $\frac{\partial p_i^*(0)}{\partial \theta_i} \geq 0$ only patterns (ii) and (iii) can arise, i.e. the equilibrium profit is either unimodal in its service level or it first declines, then increases and after reaching a local maximum proceeds to decline.

4 Simultaneous Competition

In this section, we show that under Simultaneous Competition an equilibrium exists, as long as the upper bound benchmark for the waiting time standard, $\overline{w}$, is not excessively large. Let $\underline{b} = \min_i b_i, \underline{\gamma} = \min_i \gamma_i$, $\overline{\alpha'} = \max_{i \neq j} \max_{\theta_j} \alpha'_{ij}(\theta_j)$ and $\overline{a'} = \max_i a'_i(0)$. We henceforth assume that condition (6) holds, i.e., over the feasible service level and price range, each firm maintains some market share. See §2 for a discussion and lower bounds for the intercept values $\{a_i(0)\}$, which are sufficient conditions for (6).

Theorem 3. *(Simultaneous Competition) Assume $\overline{w} \leq \sqrt[3]{\frac{4\underline{b}\underline{\gamma}}{(\overline{a'})^2}}$. There exists an equilibrium (p^*, θ^*), in the Simultaneous Competition model, with $p^{min} < p^* < p^{max}$, which satisfies the system of equations:*

$$\frac{\partial \Pi_i}{\partial p_i} = -b_i(p_i - c_i - \gamma_i) + \lambda_i = 0, \quad i = 1, \ldots, N. \tag{23}$$

$$\theta_i(p_i) = \begin{cases} \textit{the unique root of } a_i'(\theta_i)(p_i - c_i - \gamma_i) = \frac{\gamma_i}{(\overline{w}-\theta_i)^2} & , \textit{ if } \quad p_i \geq c_i + \gamma_i \left(1 + \frac{1}{\overline{w}^2 a_i'(0)}\right) \\ 0 & \textit{otherwise} . \end{cases} \tag{24}$$

Conversely, any solution of (23) and (24) is an equilibrium.

Proof. It suffices to show that the profit function Π_i is jointly concave in (p_i, θ_i). If follows from (4) and (7) that

$$\frac{\partial \Pi_i}{\partial p_i} = -b_i(p_i - c_i - \gamma_i) + \lambda_i. \tag{25}$$

$$\frac{\partial \Pi_i}{\partial \theta_i} = a_i'(\theta_i)(p_i - c_i - \gamma_i) - \frac{\gamma_i}{(\overline{w} - \theta_i)^2}. \tag{26}$$

(By (6) $\lambda_i > 0$ and λ_i equals the right hand side of (4), without the $[]^+$ - operator.) Thus, $\frac{\partial^2 \Pi_i}{\partial p_1^2} = -2b_i < 0$, $\frac{\partial^2 \Pi_i}{\partial \theta_i^2} = a_i''(\theta_i)(p_i - c_i - \gamma_i) - \frac{2\gamma_i}{(\overline{w}-\theta_i)^3} < 0$, $\frac{\partial^2 \Pi_i}{\partial \theta_i \partial p_i} = a_i'(\theta_i)$. The determinant of the Hessian is given by $-2b_i \left(a_i''(\theta_i)(p_i - c_i - \gamma_i) - \frac{2\gamma_i}{(\overline{w}-\theta_i)^3}\right) - (a_i'(\theta_i))^2 \geq 0$, provided that $\frac{4b_i\gamma_i}{\overline{w}^3} \geq (a_i'(\theta_i))^2 \iff \overline{w} \leq \min_{\theta_i} \sqrt[3]{\frac{4b_i\gamma_i}{(a_i'(\theta_i))^2}} = \sqrt[3]{\frac{4b_i\gamma_i}{(a_i'(0))^2}}$, where the last equality follows from $a_i' > 0$ and a_i' decreasing. Since $p^* = p^*(\theta^*)$, it is, by Theorem 1 in the interior of the feasible region $[p^{min}, p^{max}]$ and must therefore satisfy (23). Also, from (26), $\frac{\partial \Pi_i}{\partial \theta_i} \to -\infty$ as $\theta_i \uparrow \overline{w}$, which leaves us with the two possibilities in (24). (If $a_i'(0) \leq \frac{\gamma_i}{(\overline{w})^2(p_i - c_i - \gamma_i)}$, Π_i is decreasing in θ_i on the entire interval $[0, \overline{w}]$; otherwise, the equation $a_i'(\theta_i)(p_i - c_i - \gamma_i) = \frac{\gamma_i}{(\overline{w}-\theta_i)^2}$ has a unique root since $\lim_{\theta_i \uparrow \overline{w}} \frac{\gamma_i}{(\overline{w}-\theta_i)^2} = \infty$, and this unique root maximizes the function). □

Thus, as in the (SF) model, the only condition necessary for the existence of an equilibrium is that the upper bound benchmark for the waiting time standard fall below a specific critical value. The upper bound for $\overline{w}$ is a crude sufficient condition for the determinant of the Hessian of Π_i to be positive, and hence for Π_i to be jointly concave in (p_i, θ_i), so that the Nash-Debreu theorem can be used to guarantee the existence of an equilibrium. Alternatively, as is immediate from the proof, if service levels are measured relative to an arbitrarily large benchmark $\overline{w}$, it is sufficient that all service levels be chosen above a minimum threshold value $\underline{\theta} > 0$, or equivalently that all waiting times be chosen below a maximum value w^{max}. Condition (23) shows that in equilibrium, a firm's variable margin $p_i - c_i - \gamma_i$ is proportional to its demand volume. In particular, when all b_i coefficients are identical, a service provider is able to achieve a large demand volume if and only if it is able to obtain a large profit margin. The equilibrium conditions (23) may also be written in the form $\frac{p_i - c_i - \gamma_i}{p_i} = \frac{1}{|\epsilon^*_{ii}|}$, where ϵ^*_{ii} denotes the demand elasticity of firm i with respect to changes in its own price p_i. Thus a firm's markup expressed as a fraction of its sales price - often referred to as the *Lerner index.* See Tirole (1989) - equals, in equilibrium, the reciprocal of the absolute value of the demand elasticity. The equilibrium conditions (23) thus represent a manifestation of the *inverse elasticity rule*, noted in simpler oligopoly models. See Tirole (1989, p. 70).

As to a firm's equilibrium service level, note from (24) that it only depends on its *own* characteristics and its *own* price. Employing the implicit function theorem, one observes that a firm's equilibrium service level *increases* with its equilibrium price: for $p_i > c_i + \gamma_i \left(1 + \frac{1}{\overline{w}^2 a_i'(0)}\right)$,

$$\theta_i'(p_i) = \frac{a_i'(\theta_i)}{a_i''(\theta_i)(p_i - c_i - \gamma_i) - \frac{\gamma_i}{(\overline{w} - \theta_i)^2}} > 0, \tag{27}$$

while $\theta_i'(p_i) = 0$ for $p_i < c_i + \gamma_i \left(1 + \frac{1}{\overline{w}^2 a_i'(0)}\right)$. Moreover, for $p_i \geq c_i + \gamma_i \left(1 + \frac{1}{\overline{w}^2 a_i'(0)}\right)$, θ_i^* increases *concavely* with p_i^*, as follows directly from the second derivative of $\theta_i(\cdot)$. (24) may be used to substitute all service level variables in (23), resulting in a system of non-linear equations in the price vector p only. It is, unfortunately, not easy to solve this system directly; moreover, the possibility of multiple solutions can not be excluded a priori. In the next section we will however design a simple algorithm to compute the equilibrium price vector(s) p^* by showing that the same vector(s) is also an equilibrium in the (PF) model. Again, once the equilibrium vector p^* has been computed, the associated equilibrium service levels are immediately obtained from (24).

5 Price First Model

To analyze the two stage (PF) model, we need to start with the *second stage game* under which firms select their service level under a given and commonly known vector of prices p. We refer to the second stage game as the *service competition model.* This model is of interest, by itself, in settings in which prices are specified in a way different than through non-cooperative competition.

5.1 The Service Competition Model

Corollary 2 below shows that a *unique* equilibrium exists in the Service Competition model which arises under any given price vector p^0. This equilibrium is, in fact, given by $\theta(p^0)$ defined in (24). Moreover, the equilibrium is a *dominant* solution, i.e. $\theta_i(p^0)$ is an optimal service level choice for firm i, regardless of what choices its competitors make.

Corollary 2. *Fix a price vector* p^0. $\theta(p^0)$ *is the dominant solution in the resulting Service Competition game; moreover, a firm's equilibrium service level is independent of any of its competitors' cost or demand characteristics, their prices and the cross term functions* $\{\alpha_{ij}(\cdot)\}$. *Also, when* $\theta_i(p_i^0) > 0$, *the equilibrium service level* $\theta_i(p_i^0)$ *is increasing and concave in* p_i^0 *with* $\theta_i'(p_i^0) = \frac{-a_i'(\theta_i)}{a_i''(\theta_i)(p_i^0 - c_i - \gamma_i) - \frac{2\gamma_i}{(\overline{w}-\theta_i)^3}}$, .

Proof. The fact that $\theta(p^0)$ is a Nash equilibrium in the Service Competition game follows as a special case of Theorem 3 with the choice $p^{min} = p^{max} = p^0$. The characterization of $\theta(p^0)$ in (24) shows that the equilibrium is in fact unique and that it is a *dominant* solution, since $\theta(p^0)$ is a function of p_i, c_i and γ_i only. Finally, the monotonicity and concavity properties of $\theta_i(p^0)$ were obtained in the discussion after Theorem 3 (see (27)). □

5.2 The Price First Model: The Two Stage Game

Since, by Corollary 2, $\theta(p)$ is the *unique* equilibrium in the (*second* stage) service competition game under a given price vector p, the firms face in the *first* stage game the following reduced (equilibrium) profit functions. For all $i = 1, \dots, N$

$$\widehat{\pi}_i(p) = \Pi_i(p, \theta(p)) = \left[a_i(\theta_i(p_i)) - \sum_{j \neq i} \alpha_{ij}(\theta_j(p_j)) - b_i p_i + \sum_{j \neq i} \beta_{ij} p_j \right] (p_i - c_i - \gamma_i) - \frac{\gamma_i}{\overline{w} - \theta_i(p)} \tag{28}$$

We establish a simple one-to-one correspondence between the equilibria of the Price First competition model (PF) and those of the Simultaneous Competition model (SC). The equivalence follows from two properties of the equilibrium of the Service Competition model: first, the model has a *dominant* solution; second, the dominant choice for firm i, is independent of the price choices made by any of its competitors.

Theorem 4. *(Equivalence Between Price First and Simultaneous Competition)*

(a) *If p^* is a Nash equilibrium in the first stage game of the Price First model, then $(p^*, \theta(p^*))$ is a Nash equilibrium in the Simultaneous Competition game.*

(b) *If (p^*, θ^*) is a Nash equilibrium in the Simultaneous Competition game, then $\theta^* = \theta(p^*)$ and p^* is a Nash equilibrium in the first stage game of the Price First model.*

Proof. **(a)** Let $\theta^* = \theta(p^*)$. Assume to the contrary that for some firm i, a pair (p_i, θ_i) exists such that $\widehat{\pi}_i(p^*) = \Pi_i(p^*, \theta(p^*)) < \Pi_i(p_i, \theta_i, p^*_{-i}, \theta^*_{-i}) \leq \Pi_i(p_i, \theta_i(p_i), p^*_{-i}, \theta^*_{-i}) = \widehat{\pi}_i(p_i, p^*_{-i})$. where the last *inequality* follows from the fact that $\theta_i(p_i)$ is the optimal service level choice for firm i given the firm chooses the price p_i. Also, for $j \neq i$, $\theta^*_j = \theta_j(p^*_j)$ does not depend on firm i's choices, thus verifying the last *equality*. But $\widehat{\pi}_i(p^*) < \widehat{\pi}_i(p_i, p^*_{-i})$ contradicts that p^* is an equilibrium in the (PF) model.

(b) Note first that θ^* is an equilibrium in the service competition game under the price vector p^*. Thus, by Corollary 2, $\theta^* = \theta(p^*)$. Assume to the contrary that for some firm i, a price level p_i exists, such that $\Pi_i(p^*, \theta^*) = \Pi_i(p^*, \theta(p^*)) = \widehat{\pi}_i(p^*) < \widehat{\pi}_i(p_i, p^*_{-i}) = \Pi_i(p_i, \theta_i(p_i), p^*_{-i}, \theta(p^*)_{-i}) = \Pi_i(p_i, \theta(p_i), p^*_{-i}, \theta^*_{-i})$. Thus, in the (SC) game, firm i can profitably deviate from the equilibrium by selecting the pair $(p_i, \theta_i(p_i))$, a contradiction.

□

Theorems 3 and 4 establish the existence of an equilibrium in the (PF) model.

Corollary 3. *Let $\overline{w} \leq \sqrt[3]{\frac{4b\gamma}{(a')^2}}$. There exists a Nash equilibrium $p^{min} < p^* < p^{max}$ for the (first stage game of the) Price First model.*

Corollary 3 does not guarantee that the equilibrium is unique. In view of Theorem 4, all we know is that an equilibrium p^* must satisfy the system of N non-linear equations that results after substituting in (24) all variables θ_i by the functions $\theta_i(p_i)$. As mentioned in §4, it is not apparent how this system is to be solved directly. However, the next theorem states that the first stage game is supermodular (under a slightly different upper bound for $\overline{w}$), so that p^* can, again, be computed by a tatônnement scheme.

Theorem 5. *(Price First Model. Characterization and Computation of Equilibria). Assume $\overline{w} \leq \sqrt[3]{\frac{2\beta\gamma}{a'\alpha'}}$.*

(a) *The first stage game in the Price First model is supermodular.*

(b) *The set of equilibria is a sublattice of* $\mathbb{R}^N$ *and in particular has a component-wise largest and a component-wise smallest element* $\overline{p}, \underline{p}$ *respectively.*

(c) *The tatônnement scheme converges to a Nash equilibrium. When started with* $p^0 = p^{max}(p^{min})$ *this scheme generates a decreasing (increasing) sequence of price vectors converging to* $\overline{p}(\underline{p})$.

Proof. (a) In view of (28), to show that $\widehat{\pi}_i$ has the supermodularity property, it suffices to show that $[\beta_{ij}p_j - \alpha_{ij}(\theta_j(p_j))]\, p_i$ is a supermodular function, which holds if and only if the function $\phi_{ij}(p_j) = [\beta_{ij}p_j - \alpha_{ij}(\theta_j(p_j))]$ is *increasing* in p_j. (Note that all other terms in (28) depend on a *single* price variable only). For $p_j < c_j + \gamma_j\left(1 + \frac{1}{\overline{w}^2 a_j'(0)}\right)$, $\phi_{ij}(p_j)$ is increasing because $\theta_j(p_j) = 0$. Moreover, $\phi_j(p_j)$ is continuous everywhere, and for $p_j > c_j + \gamma_j\left(1 + \frac{1}{\overline{w}^2 a_j'(0)}\right)$, $\phi_{ij}(p_j)$ is differentiable with $\phi_{ij}'(p_j) = \beta_{ij} + \frac{\alpha_{ij}'(\theta_j)a_j'(\theta_j)}{a_j''(\theta_j)(p_j - c_j - \gamma_j) - \frac{2\gamma_j}{(\overline{w}-\theta_j)^3}} \geq \beta_{ij} - \frac{\alpha_{ij}'(\theta_j)a_j'(\theta_j)(\overline{w}-\theta_j)^3}{2\gamma_j} \geq \beta_{ij} - \frac{\alpha_{ij}'(\theta_j)a_j'(\theta_j)\overline{w}^3}{2\gamma_j} \geq 0$, where the first inequality follows from the concavity of $a_j(\cdot)$ and the last inequality from the bound for $\overline{w}$. (b) and (c): Follow from the supermodularity of the first stage game. See Topkis (1998). □

The tatônnement scheme reduces to the repeated optimization of the single variable functions $\widehat{\pi}_i^*(\cdot|p_{-i}^k)$, see (28). This remains somewhat complex, as these profit functions in general fail to be concave, and in addition fail to be given in closed form, since the functions $\theta_i(p_i)$ are not. In §6 we design an alternative and much simpler scheme (based on Theorem 4's equivalence result) which converges to the equilibria of the (PF) model.

The following Round Robin scheme provides an alternative iterative method that converges monotonically to $\overline{p}$ ($\underline{p}$) when started at p^{max} (p^{min}): Traversing the firms in a fixed Round Robin permutation, each firm selects a best response price to the prevailing price choices of his competitors. Note that the tatônnement (or Round Robin) scheme can be used to numerically verify whether multiple equilibria exist; the equilibrium is unique if and only if the scheme converges to the same limit when started at p^{max} and p^{min}. Indeed, this test has always been met, i.e. a unique equilibrium does exist in each of thousands of numerical instances we have evaluated. Furthermore, even if multiple equilibria were to arise in some (yet undiscovered) instances, it can be shown that the component-wise largest equilibrium $\overline{p}$ is preferred by all firms, provided the upper bound benchmark $\overline{w}$ is bounded by $\overline{w} \leq \sqrt[3]{\frac{4\beta\gamma}{\underline{\underline{a'\alpha'}}}}$, a bound similar to the one assumed in Theorem 5. Verification of this statement follows a general argument in Theorem 7 of Milgrom and Roberts (1990): If $p^1 \geq p^2$ are a pair of Nash equilibria, $\widehat{\pi}_i(p^1) \geq \widehat{\pi}_i(p_i^2, p_{-i}^1) \geq \widehat{\pi}_i(p_i^2, p_{-i}^2) = \widehat{\pi}_i(p^2)$. The first inequality follows from the fact that p^1 is a Nash equilibrium. The second inequality follows from the fact that $\widehat{\pi}_i$ is increasing in p_j, see the proof of Theorem 5(c).

The equivalence between the Price First model and the Simultaneous Competition model, established in Theorem 4, also allows us to characterize the equilibria of the latter.

Corollary 4. *Assume $\overline{w} \leq \sqrt[3]{\frac{2\beta\gamma}{\underline{\underline{a'\alpha'}}}}$. The set of equilibria in the Simultaneous Competition model contains a component-wise smallest pair $(\underline{p}, \underline{\theta})$ and a component-wise largest pair $(\overline{p}, \overline{\theta})$, i.e., for any equilibrium (p^*, θ^*): $\underline{p}_i \leq p_i^* \leq \overline{p}_i$; $\underline{\theta}_i \leq \theta_i^* \leq \overline{\theta}_i$, $i = 1, \ldots, N$*

Proof. Let $\overline{p}(\underline{p})$ denote the component-wise largest (smallest) equilibrium in the (PF) model, which exists by Theorem 5. Define $\overline{\theta} = \theta(\overline{p})$ and $\underline{\theta} = \theta(\underline{p})$. By Theorem 4, $(\overline{p}, \overline{\theta})$ and $(\underline{p}, \underline{\theta})$ are Nash equilibria in the (SC) model. Consider any other equilibrium pair (p^*, θ^*) of this model. Again by Theorem 4, $\theta^* = \theta(p^*)$, and p^* is an equilibrium of the (PF) model with $\underline{p} \leq p^* \leq \overline{p}$. Finally, the inequalities $\underline{\theta} \leq \theta^* \leq \overline{\theta}$ follow from $\underline{\theta} = \theta(\underline{p}), \theta^* = \theta(p^*), \overline{\theta} = \theta(\overline{p})$ and the monotonicity of the $\theta(\cdot)$ vector function , see (24). □

6 A Comparison Of Equilibria In The Three Competition Models

As shown in Theorem 4, the PF and SC models share the same set of equilibria. Thus, prior knowledge of the firms' prices has no impact on their equilibrium service level choices. The same fails to be true, however, when comparing the equilibria in the (SF) model with those in the other two competition models. In this §, we show that under a variant of condition (22), the (SF) model results in higher prices, higher service levels and higher demand volumes for *all* firms.

Thus, if firms make their strategic decisions *sequentially*, selecting service levels, hence waiting time standards *first*, this results in an equilibrium with *higher* service levels, prices and demand volumes, as compared to the equilibrium reached in the (SC) model. This phenomenon bears close similarity to the "fat cat" effect, a term coined by Fudenberg and Tirole (1984). Like a "fat cat", the firms are inclined to "overinvest" in service and capacity in order to deter the competitors in the subsequent price competition. Interestingly, the same phenomenon fails to occur in the (PF) model; that is, when competitors learn up-front about the firms' price choices, this does *not* provide an incentive to either "underprice" or "overprice" compared to the (SC) model. Instead, the exact same equilibrium arises.

Theorem 5 and Corollary 1 show that both the (PF) and the (SF) model have a component-wise smallest equilibrium and that these equilibria arise as the limit of a tatônnement scheme, started with $p^0 = p^*(0)$ and $\theta^0 = 0$, respectively. To establish the above ranking of the equilibria in the two models we show that in each iteration, the tatônnement scheme for the (PF) model generates a price and associated service level vector which is component-wise smaller than the price and service level vector generated by the tatônnement scheme for the (SF) model.

Let $p^k_{i,PF}$ and $\theta^k_{i,PF} = \theta_i(p^k_{i,PF})$ denote firm i's price and service level, generated in the k-th iteration of the tatônnement scheme for the (PF) model with $p^0_{PF} = p^*(0), i = 1, \ldots, N$. As explained in §5, it is somewhat cumbersome to determine p^{k+1}_{PF} from p^k_{PF} directly by computing for each $i = 1, \ldots, N$ the maximum of the function $\widehat{\pi}_i(\cdot|p^k_{-i,PF})$. We first show that the sequence $\{p^k_{PF}\}$ can be generated via a simpler iterative scheme:

Lemma 2. *Assume (22). Let $\theta^0_{PF} = 0$ and $p^0_{PF} = p^*(0)$. Consider the iterative scheme which starts at $\theta^0 = \theta^0_{PF}$ and $p^0 = p^0_{PF}$ and in the $k+1$-st iteration ($k \geq 0$) generates the vectors θ^{k+1} and p^{k+1} as follows:*

$$\theta^{k+1}_i = \arg\max_{\theta_i} \widehat{\Pi}^{k+1}_i(\theta_i|\theta^k_{-i}) = \arg\max_{\theta_i} \left\{ \frac{1}{4b_i} \left(a_i(\theta_i) - \sum_{j \neq i} \alpha_{ij}(\theta^k_j) + \sum_{j \neq i} \beta_{ij} p^k_j - \kappa_i \right)^2 - \frac{\gamma_i}{\overline{w} - \theta_i} \right\}$$

$$p^{k+1}_i = \frac{a_i(\theta^{k+1}_i) - \sum_{j \neq i} \alpha_{ij}(\theta^k_j) + \sum_{j \neq i} \beta_{ij} p^k_j + \kappa_i}{2b_i}$$

Then, $p^0 = p^0_{PF} \leq p^1 = p^1_{PF} \leq p^2 = p^2_{PF} \leq \cdots \uparrow \underline{p}_{PF}$, $0 = \theta^0 \leq \theta^1 = \theta^1_{PF} \leq \theta^2 = \theta^2_{PF} \leq \cdots \uparrow \underline{\theta}_{PF} \equiv \theta(\underline{p}_{PF})$

Proof. By Theorem 5(c), $\{p^k_{PF}\} \uparrow \underline{p}_{PF}$ since $p^0_{PF} = p^*(0) \leq \underline{p}_{PF}$. To verify the latter inequality, note from Theorem 4 that $(\underline{p}_{PF}, \theta(\underline{p}_{PF}))$ is an equilibrium in the Simultaneous

Competition model so that $\underline{p}_{PF}$ is the price equilibrium in the price competition game which arises under the fixed service level vector $\theta(\underline{p}_{PF})$. Thus $\underline{p}_{PF} = p^*(\underline{\theta}_{PF}) \geq p^*(0)$, by (22). We now show, by induction, that for all $k \geq 0$

$$\begin{aligned} p^{min} \leq \quad p^0 &= p^0_{PF} \leq p^1 = p^1_{PF} \leq \cdots \leq p^k = p^k_{PF} \leq p^u \\ \theta^0 &= \theta^0_{PF} \leq \theta^1 = \theta^1_{PF} \leq \cdots \leq \theta^k = \theta^k_{PF} \leq [\overline{w}, \overline{w}, \ldots, \overline{w}]^T. \end{aligned} \tag{29}$$

By Theorem 1(a), (29) clearly holds for $k = 0$; assume it holds for some $k \geq 0$. To show that it holds for $k+1$ as well, note from Theorem 5(c) that $p^k_{PF} \leq p^{k+1}_{PF}$ and $\theta^k_{PF} = \theta(p^k_{PF}) \leq \theta(p^{k+1}_{PF}) = \theta^{k+1}_{PF}$ where the inequality follows from the fact that the $\theta(\cdot)$ function is increasing. It thus suffices to show that

$$p^{k+1} = p^{k+1}_{PF}; \theta^{k+1} = \theta^{k+1}_{PF}. \tag{30}$$

Per definition,

$$\begin{aligned} p^{k+1}_{i,PF} = \arg\max_{p_i} \widehat{\pi}_i(p_i, p^k_{-i,PF}) &= \arg\max_{p_i} \Pi_i(p_i, \theta_i(p_i), p^k_{-i,PF}, \theta(p^k_{-i,PF})) \\ &= \arg\max_{p_i} \Pi_i(p_i, \theta_i(p_i), p^k_{-i}, \theta(p^k_{-i})) \end{aligned} \tag{31}$$

where the second equality follows from (28) and the last one from the induction assumption. Therefore

$$(p^{k+1}_{i,PF}, \theta^{k+1}_{i,PF}) = (p^{k+1}_{i,PF}, \theta_i(p^{k+1}_{i,PF})) = \arg\max_{p_i,\theta_i} \Pi_i(p_i, \theta_i, p^k_{-i}, \theta(p^k_{-i})) \tag{32}$$

The first equality follows from the definition of $\theta^{k+1}_{i,PF}$ and the second one from

$$\max_{p_i}\left\{\max_{\theta_i} \Pi_i(p_i, \theta_i, p^k_{-i}, \theta^k_{-i})\right\} = \max_{p_i,\theta_i} \Pi_i(p_i, \theta_i, p^k_{-i}, \theta(p^k_{-i})) = \max_{\theta_i}\left\{\max_{p_i} \Pi_i(p_i, \theta_i, p^k_{-i}, \theta^k_{-i})\right\} \tag{33}$$

and the fact that $p^{k+1}_{i,PF}$ is the largest maximizer of the function to the left of (31). Thus, to establish (30) and hence to complete the induction step it suffices to show that the pair $(p^{k+1}_i, \theta^{k+1}_i)$ defined in the Lemma represents the largest maximizer of the function to the right of (33). Observe that the function within curled brackets to the right of (33) is a quadratic function in p_i, i.e. the unconstrained maximizer of this function is

$$p^{k+1}_i(\theta_i) \equiv \frac{a_i(\theta_i) - \sum_{j \neq i} \alpha_{ij}(\theta^k_j) + \sum_{j \neq i} \beta_{ij} p^k_j + \kappa_i}{2b_i}, i = 1, \ldots, N. \tag{34}$$

and since

$$
\begin{aligned}
p_i^{min} = c_i + \gamma_i \quad &\leq \quad \frac{\overline{a}_i^{min} + \sum_{j \neq i} \beta_{ij}(c_j + \gamma_j) + \kappa_i}{2b_i} \\
&\leq \quad \frac{a_i(\theta_i) - \sum_{j \neq i} \alpha_{ij}(\theta_j^k) + \sum_{j \neq i} \beta_{ij} p_j^k + \kappa_i}{2b_i} \\
&= \quad p_i^{k+1}(\theta_i) \leq \frac{a_i(\overline{w}) + \sum_{j \neq i} \beta_{ij} p_j^u + \kappa_i}{2b_i} = p_i^u \leq p_i^{max}
\end{aligned} \tag{35}
$$

it is its constrained maximizer as well. (The first and last inequalities follow from (5) and (17) and the second one from the definition of $\overline{a}_i^{min}$ and $p_j^{min} \leq p_j^k$, by the induction assumption. The third inequality follows from the monotonicity of $a(\cdot)$ and $p_j^k \leq p_j^u$ by the induction assumption. The last equality follows from the fact that p^u is the unique solution of the equation $Ax = a(\overline{w}) + \kappa$.) This implies that the value of θ_i which is the largest maximizer of the function to the right of (33) is also the largest maximizer of

$$
\begin{aligned}
&\left(a_i(\theta_i) - \sum_{j \neq i} \alpha_{ij}(\theta_j^k) - b_i p_i^{k+1}(\theta_i) + \sum_{j \neq i} \beta_{ij} p_j^k \right) \left(p_i^{k+1}(\theta_i) - c_i - \gamma_i \right) - \frac{\gamma_i}{\overline{w} - \theta_i} \\
= \quad &\frac{1}{4} \left(a_i(\theta_i) - \sum_{j \neq i} \alpha_{ij}(\theta_j^k) + \sum_{j \neq i} \beta_{ij} p_j^k - \kappa_i \right) \left(\frac{a_i(\theta_i) - \sum_{j \neq i} \alpha_{ij}(\theta_j^k) + \sum_{j \neq i} \beta_{ij} p_j^k - \kappa_i}{b_i} \right) - \frac{\gamma_i}{\overline{w} - \theta_i} \\
= \quad &\frac{1}{4b_i} \left(a_i(\theta_i) - \sum_{j \neq i} \alpha_{ij}(\theta_j^k) + \sum_{j \neq i} \beta_{ij} p_j^k - \kappa_i \right)^2 - \frac{\gamma_i}{\overline{w} - \theta_i}.
\end{aligned}
$$

In other words, the values $(p_i^{k+1}, \theta_i^{k+1})$ defined in the Lemma represent the largest maximizers of the function to the right of (33), and as argued, must coincide with $(p_{i,PF}^{k+1}, \theta_{i,PF}^{k+1})$, thus verifying (30). □

In conclusion, to compute the equilibria in the (PF) model (and hence in the (SC) model), it is considerably easier to employ the scheme of Lemma 2 as opposed to the basic tatônnement scheme applied to the (PF) model. While this scheme continues to require that in each iteration for each firm, a non-linear single variable function be maximized, at least this function is now given in a simple closed form. In the important special case in which the $a_i(\cdot)$ functions are affine, the roots of the first derivative of each of the functions $\widehat{\Pi}_i^{k+1}(\theta_i|\cdot)$, hence its local optima, can be found by solving a cubic equation. (The roots of a cubic equation can be computed in closed form.) Finally, it is easily verified that the Lemma applies to the fully general model, (without condition (22)) as long as one chooses to start at $p^0 = p_{PF}^0 = p^{min}$.

We now establish the ranking of the equilibrium in the (SF) model relative to that in the other two models. Let $\theta^k_{i,SF}$ and $p^k_{i,SF} = p^*_i(\theta^k_{SF})$ denote firm i's price and service level generated in the k-th iteration of the tatônnement scheme for the (SF) model, with $\theta^0_{i,SF} = 0, i = 1, \ldots, N$. We need a slightly stronger version of condition (22) which maintains the inequalities for $i \neq j$, but restricts those for $i = j$:

$$\begin{aligned} (A^{-1})_{ij} a'_j(\theta_j) &\geq \sum_{l \neq j} (A^{-1})_{il} \alpha'_{lj}(\theta_j), \quad \forall i \neq j = 1, \ldots, N \\ \left(1 - \frac{1}{2\delta_i}\right)(A^{-1})_{ii} a'_i(\theta_i) &\geq \sum_{l \neq i} (A^{-1})_{il} \alpha'_{li}(\theta_i), \quad \forall i = 1, \ldots, N \end{aligned} \tag{36}$$

Note from Lemma 1 that $0 \leq 1 - \frac{1}{2\delta_i} \leq \frac{1}{2}$.

Theorem 6. *(Comparison Between Equilibria) Assume (36) holds. The equilibrium pair* $(p^*(\underline{\theta}_{SF}), \underline{\theta}_{SF})$ *in the Service Level First model is component-wise at least as large as the pair* $(\underline{p}_{PF}, \underline{\theta}_{PF})$. *In particular, if the Simultaneous Competition model and hence the Price First model, has a unique equilibrium, all equilibria of the Service Level First model are component-wise at least as large.*

Proof. We show by induction that

$$\theta^k_{SF} \geq \theta^k_{PF}, k = 0, 1, \ldots \tag{37}$$

$$p^*(\theta^k_{PF}) \geq p^k_{PF}, k = 1, \ldots \tag{38}$$

By (37) and the monotonicity of $p^*(\cdot)$ under (36), hence (22), this establishes $p^k_{SF} = p^*(\theta^k_{SF}) \geq p^*(\theta^k_{PF}) \geq p^k_{PF}$. The theorem follows since the sequences $\{(p^k_{PF}, \theta^k_{PF})\}$ and $\{(p^*(\theta^k_{SF}), \theta^k_{SF})\}$ converge to $(\underline{p}_{PF}, \underline{\theta}_{PF})$ and $(p^*(\underline{\theta}_{SF}), \underline{\theta}_{SF})$ respectively.

Note first that the sequences $\{(p^k_{PF}, \theta^k_{PF})\}$ and $\{(p^*(\theta^k_{SF}), \theta^k_{SF})\}$ are component-wise increasing. The monotonicity of the former sequence follows from Lemma 2, that of $\{\theta^k_{SF}\}$ from Corollary 1(b) and the monotonicity of $\{p^*(\theta^k_{SF})\}$ then follows from (22). To prove (37) and (38), note that the inequalities hold for $k = 0$. Assume they hold for some arbitrary $k \geq 0$. $\theta^{k+1}_{i,SF}$ is the *largest* maximizer of the function $\pi^*(\theta_i|\theta^k_{-i,SF})$. Similarly $\theta^{k+1}_{i,PF}$ is the largest maximizer of the function $\widehat{\Pi}^{k+1}_i(\theta_i|\theta^k_{-i,PF})$. Thus, to show that $\theta^{k+1}_{SF} \geq \theta^{k+1}_{PF}$, it suffices to show that for all i

$$\frac{\partial \pi^*_i(\theta_i|\theta^k_{-i,SF})}{\partial \theta_i} \geq \frac{\partial \widehat{\Pi}^{k+1}_i(\theta_i|\theta^k_{-i,PF})}{\partial \theta_i}, \quad \forall \theta_i \geq \theta^k_{i,SF} \tag{39}$$

To show the sufficiency of (39), recall that $\theta_{SF}^{k+1} \geq \theta_{SF}^k$. For any $\theta_{i,PF}^{k+1} \geq \theta \geq \theta_{i,SF}^k$, $\pi_i^*(\theta_{i,PF}^{k+1}, \theta_{-i,SF}^k) - \pi_i^*(\theta, \theta_{-i,SF}^k) = \int_\theta^{\theta_{i,PF}^{k+1}} \frac{\partial \pi_i^*(x, \theta_{-i,SF}^k)}{\partial \theta_i} dx \geq \int_\theta^{\theta_{i,PF}^{k+1}} \frac{\partial \widehat{\Pi}_i^{k+1}(x|\theta_{-i,PF}^k)}{\partial \theta_i} dx = \widehat{\Pi}_i^{k+1}(\theta_{i,PF}^{k+1}|\theta_{-i,PF}^k) - \widehat{\Pi}_i^{k+1}(\theta|\theta_{-i,PF}^k) \geq 0$, since $\theta_{i,PF}^{k+1}$ is a global maximizer of $\widehat{\Pi}_i^{k+1}$. The largest global maximizer of $\pi_i^*(\theta_i, \theta_{-i,SF}^k)$ must thus be larger than or equal to $\theta_{i,PF}^{k+1}$.

In view of (16) and Lemma 4, (39) is equivalent to $2b_i \frac{\partial p_i^*(\theta_i, \theta_{-i,SF}^k)}{\partial \theta_i}(p_i^*(\theta_i, \theta_{-i,SF}^k) - c_i - \gamma_i) \geq \frac{a_i'(\theta_i)}{2b_i}\left(a_i(\theta_i) - \sum_{j \neq i} \alpha_{ij}(\theta_{j,PF}^k) + \sum_{j \neq i} \beta_{ij} p_{j,PF}^k - \kappa_i\right)$. Using Theorem 1(a) and (22) we obtain that (39) is equivalent to

$$
\begin{aligned}
& 2b_i\left((A^{-1})_{ii}a_i'(\theta_i) - \sum_{j\neq i}(A^{-1})_{ij}\alpha_{ji}'(\theta_i)\right) \\
& \left(\frac{a_i(\theta_i) - \sum_{j\neq i}\alpha_{ij}(\theta_{j,SF}^k) + \sum_{j\neq i}\beta_{ij}p_j^*(\theta_i, \theta_{-i,SF}^k) + \kappa_i}{2b_i} - c_i - \gamma_i\right) \\
= & \left((A^{-1})_{ii}a_i'(\theta_i) - \sum_{j\neq i}(A^{-1})_{ij}\alpha_{ji}'(\theta_i)\right) \\
& \left(a_i(\theta_i) - \sum_{j\neq i}\alpha_{ij}(\theta_{j,SF}^k) + \sum_{j\neq i}\beta_{ij}p_j^*(\theta_i, \theta_{-i,SF}^k) + \kappa_i\right) \\
\geq & \frac{a_i'(\theta_i)}{2b_i}\left(a_i(\theta_i) - \sum_{j\neq i}\alpha_{ij}(\theta_{j,PF}^k) + \sum_{j\neq i}\beta_{ij}p_{j,PF}^k - \kappa_i\right), \forall \theta_i \geq \theta_{i,SF}^k
\end{aligned} \tag{40}
$$

Both sides of the inequality are given as a product of two factors. Since the factors to the left of the inequality are easily seen to be positive, as is the first factor to its right, it suffices to show that the first (second) factor to the left dominates the first (second) factor to its right:

$$(A^{-1})_{ii}a_i'(\theta_i) - \sum_{j\neq i}(A^{-1})_{ij}\alpha_{ji}'(\theta_i) \geq \frac{a_i'(\theta_i)}{2b_i}, \forall \theta_i \geq \theta_{i,SF}^k \tag{41}$$

$$\sum_{j\neq i}\beta_{ij}p_j^*(\theta_i, \theta_{-i,SF}^k) - \sum_{j\neq i}\alpha_{ij}(\theta_{j,SF}^k) \geq \sum_{j\neq i}\beta_{ij}p_{j,PF}^k - \sum_{j\neq i}\alpha_{ij}(\theta_{j,PF}^k), \forall \theta_i \geq \theta_{i,SF}^k \tag{42}$$

(41) coincides with the second inequality in (36), while $\sum_{j\neq i}\beta_{ij}p_j^*(\theta_i, \theta_{-i,SF}^k) - \sum_{j\neq i}\alpha_{ij}(\theta_{j,SF}^k) = \lambda_i^*(\theta_i, \theta_{-i,SF}^k) + b_i p_i^*(\theta_i, \theta_{-i,SF}^k) \geq \lambda_i^*(\theta_{SF}^k) + b_i p_i^*(\theta_{SF}^k) \geq \lambda_i^*(\theta_{PF}^k) + b_i p_i^*(\theta_{PF}^k) = \sum_{j\neq i}\beta_{ij}p_j^*(\theta_{PF}^k) - \sum_{j\neq i}\alpha_{ij}(\theta_{j,PF}^k) \geq \sum_{j\neq i}\beta_{ij}p_{j,PF}^k - \sum_{j\neq i}\alpha_{ij}(\theta_{j,PF}^k)$ verifies (42). (Here $\lambda^*(\theta)$ is defined as in Theorem 1 to denote the vector of equilibrium demand volumes in the price competition game under the service level θ. Both equalities are immediate from (4). The first two inequalities follow from the fact that under (36), and

hence (22), $\lambda^*(\cdot)$ and $p^*(\cdot)$ are increasing while $\theta_i \geq \theta^k_{i,SF}$ and $\theta^k_{i,SF} \geq \theta^k_{i,PF}$, the latter by the induction assumption. Finally, the last inequality is immediate from the induction assumption as well.)

This completes the proof that $\theta^{k+1}_{SF} \geq \theta^{k+1}_{PF}$. To complete the induction proof, note that $p^*_i(\theta^{k+1}_{PF}) = \frac{a_i(\theta^{k+1}_{i,PF}) - \sum_{j\neq i} \alpha_{ij}(\theta^{k+1}_{j,PF}) + \sum_{j\neq i} \beta_{ij} p^*_j(\theta^{k+1}_{PF}) + \kappa_i}{2b_i} \geq \frac{a_i(\theta^{k+1}_{i,PF}) - \sum_{j\neq i} \alpha_{ij}(\theta^{k}_{j,PF}) + \sum_{j\neq i} \beta_{ij} p^*_j(\theta^{k}_{PF}) + \kappa_i}{2b_i} \geq \frac{a_i(\theta^{k+1}_{i,PF}) - \sum_{j\neq i} \alpha_{ij}(\theta^{k}_{j,PF}) + \sum_{j\neq i} \beta_{ij} p^{k}_{j,PF} + \kappa_i}{2b_i} = p^{k+1}_{i,PF}$ so that (38) holds for $k+1$ as well. (The first identity follows from Theorem 1(a). The first inequality can be verified in the same way as (42) since $\theta^{k+1}_{PF} \geq \theta^k_{PF}$. The second inequality follows from the induction assumption, and the last identity from Lemma 4). □

Theorem 6 ranks the component-wise smallest equilibria in the various competition models, and is therefore somewhat inconclusive with respect to other equilibria in case the (SC) model fails to have a unique equilibrium. However, in *all* instances evaluated in our numerical study, the (SC) model *has* a *unique* equilibrium and so does the (SF) model; the former equilibrium is therefore indeed component-wise smaller than the latter. The tatônnement scheme can be viewed as a possible dynamic adjustment process by which the firms adapt their choices and converge to an equilibrium, in addition to it serving as an efficient algorithm for its computation. Indeed, as stated in Vives (page 49), "although this adjustment process can (and has) been criticized for being ad hoc, it can also be interpreted as a crude way of expressing the *bounded rationality* of agents". The proof of Theorem 6 shows that under this dynamic adjustment process, the service levels and prices adopted under the (SF) setting are larger than those under (SC) and (PF) competition, at *each stage* of the adjustment process, and not just in equilibrium.

While we have argued that condition (22), hence (36) are likely to hold, it may sometimes be violated. The following example shows that the ranking between the equilibria in the three competition models, as specified by Theorem 6, may fail to apply when condition (36) is violated.

Example 2: As in Example 1, consider an industry with $N = 3$ firms, $\overline{w} = 100$, and cost parameters $c_1 = c_2 = 20, c_3 = 5$, while $\gamma_1 = \gamma_2 = 20, \gamma_3 = 35$. The example may, therefore, once again, apply to a setting with firm 3 an established local service provider and firms 1 and 2 competitors that have entered the local market more recently from a foreign or remote location, where capacity costs (γ) are lower, but the per customer access costs (c) are higher. In *this* example firms experience identical price sensitivities,i.e. $b_i = 10$ and $\beta_{ij} = 4.5, \forall i \neq j$. Finally, $\overline{a}_1(\theta) = 145 + 0.1\theta_1 - \epsilon\theta_2 - \eta\theta_3; \overline{a}_2(\theta) = 145 + 0.1\theta_2 - \epsilon\theta_1 - \eta\theta_3; \overline{a}_3(\theta) = 235 + 0.1\theta_3 - \epsilon\theta_2 - \epsilon\theta_1$ Thus, as in Example 1, firms 1 and 2 have identical characteristics and the more established firm 3 captures under identical prices and service levels a larger demand volume ($\overline{a}_3(\theta) - \overline{a}_1(\theta) = \overline{a}_3(\theta) - \overline{a}_2(\theta) = 90$). In Table 1 we evaluate 6 instances by combining three values for η, ($\eta = 0, 0.01, 0.05$) with two values for ϵ, ($\epsilon = 0, 0.04$), referring to the case $\eta = \epsilon = 0$ as the *base case.* Since firms 1 and 2 are identical, we report equilibrium prices, demand volumes, waiting times and profits for firms 1 and 3, under the (SF) and (PF) models. In all instances, both the (PF) and (SF) models have a unique equilibrium, since the respective tatônnement schemes converge to the same limits irrespective of their starting points.

Condition (22) is satisfied in all but the last instance, in which the slopes of the cross term functions are so large that, an exclusive service level improvement by firm 1 or firm 2 results in an increase of this firm's demand volume by an amount ten times as large as when the service level improvement occurs on an industry-wide basis. (Since all $a_i(\cdot)$, $\alpha_{ij}(\cdot)$ functions are affine, both sides of the inequalities of (22) are constants; the matrix A^{-1} has $A^{-1}_{ii} = 0.0575, A^{-1}_{ij} = 0.0169$ for $i \neq j$). The stricter condition (36) is *only* satisfied in the first two instances. In accordance with Theorem 6, the (SF) competition model results in higher prices and service levels for all firms, in both cases. At the same time, violations of this ranking arise in the remaining 4 instances in which (36) is violated, even though all equilibrium *prices* are higher under (SF) competition. In the third (fourth and fifth) instance(s), firm 3 (firms 1 and 2) offers (offer) a lower service level while firms 1 and 2 (3) offer(s) a higher service level under (SF) as compared to the two other types of competition. In the *last* instance, all firms offer a lower service level under (SF). One might expect that the larger the cross terms in the intercept functions the larger the competitive pressure to improve service. The instances in Table 1 show however, that the opposite may occur. As either ϵ or η is increased, the equilibrium prices *increase* under (SF) competition, but they *decrease* under (PF).

Table 1: **Equilibria Under Different Cross Term Functions**

ϵ, η	Type	(22)	(36)	p_1	w_1	λ_1	π_1	p_3	w_3	λ_3	π_3
0,0	SF	✓	✓	65.55	5.18	105.46	1105.48	69.22	5.38	142.19	2012.48
	SC			65.54	5.76	105.41	1105.06	69.21	5.93	142.14	2011.89
0,0.01	SF	✓	✓	65.62	5.19	106.16	1120.25	69.25	5.55	142.49	2021.46
	SC			65.47	5.78	105.65	1100.23	69.18	5.94	141.82	2002.95
0,0.05	SF	✓		65.89	5.22	108.91	1179.53	69.37	6.43	143.69	2056.90
	SC			65.19	5.86	106.61	1080.58	69.06	5.96	140.56	1967.40
0.04,0	SF	✓		65.94	5.99	109.45	1192.05	69.77	5.43	147.74	2173.53
	SC			65.13	5.88	105.11	1059.35	68.65	6.05	144.07	1958.85
0.04,0.01	SF	✓		66.01	6.00	110.15	1207.36	69.80	5.60	148.05	2182.82
	SC			65.07	5.90	105.35	1054.46	68.62	6.06	143.75	1950.05
0.04,0.05	SF			66.29	6.03	112.90	1268.74	69.92	6.48	149.24	2219.48
	SC			64.79	5.98	106.32	1034.56	68.50	6.09	142.49	1915.05

7 Numerical Investigations and Generalizations

It is of interest to investigate how the equilibrium behavior in the various competition models is affected by the number of firms N. We illustrate this for the SC-model, in the special case where all $a(\cdot)$ and $\alpha(\cdot)$ functions are affine and the model is symmetric, i.e., for all $i = 1, \ldots, N$, and given constants a^0, a, b, α and β

$$\lambda_i = a^0 + a\theta_i - \sum_{j \neq i} \alpha\theta_j - bp_i + \sum_{j \neq i} \beta p_j, i = 1, \ldots, N \tag{43}$$

Moreover, for all $i = 1, \ldots, N$, $c_i = c, \gamma_i = \gamma, p_i^{max} = p^{max}$ for given constants c, γ, and p^{max}. Let $\rho \equiv \frac{(N-1)\beta}{b}$ and $\sigma \equiv \frac{(N-1)\alpha}{a}$. By (D), $0 < \rho \leq 1$. Similarly, no firm experiences a reduction in its demand volume if all firms increase their service level by the same amount, i.e. $0 < \sigma \leq 1$.

It follows from Theorem 3, that any solution of (23) and (24) is an equilibrium, and that a solution to (23) and (24) exist. This system of equations has a symmetric solution $p_1^* = \cdots = p_N^* = p^{eq}$ and $\theta_1^* = \cdots = \theta_N^* = \theta^{eq}$, where p^{eq} and $w^{eq} = \overline{w} - \theta^{eq}$ satisfy

$$p^{eq} = \frac{a^0 + a(1-\sigma)(\overline{w} - w^{eq}) + b(c+\gamma)}{b(2-\rho)} \tag{44}$$

$$w^{eq} = \begin{cases} \text{the unique root on } [0, \overline{w}] \text{ of } C^*(w) \equiv w^3 - w^2 \frac{a^0 + a(1-\sigma)\overline{w} - (c+\gamma)b(1-\rho)}{a(1-\sigma)} + \frac{\gamma b(2-\rho)}{a^2(1-\sigma)}, \\ \qquad if \gamma \leq \frac{a^2(1-\sigma)}{b(2-\rho)} (\overline{w})^2 \left[\frac{a^0 - (c+\gamma)b(1-\rho)}{a(1-\sigma)} - \overline{w} \right] \\ \overline{w}, \qquad \text{otherwise} \end{cases} \tag{45}$$

((44) follows from (23) by substituting $p_i = p^{eq}$ and $\theta_i = \overline{w} - w^{eq}$. Substituting the same identities, as well as (44) into (24), we obtain (45).) Assume, in addition, that $2b > a + \sigma a + \rho b \Leftrightarrow (2-\rho)b > (1-\sigma)a$, so that the $(2N \times 2N)$ Jacobian of the first order conditions (23) and (24) is negative semi definite, since each diagonal element is negative while its absolute value exceeds (dominates) the sum of the absolute values of the off-diagonal elements in its row. Thus, by the Gale - Nikaido theorem, the equilibrium described by (44) and (45) is unique. (See, e.g. Vives (2000).)

How the equilibrium varies with N, depends heavily on how the parameters in the demand equations (43) depend on N. Consider the case where the direct price and service level sensitivities, a and b, are independent of N, i.e., $a(N) = a$ and $b(N) = b$, for given constants $a > 0$ and $b > 0$, while the same applies to ρ and σ. (This means that $\beta(N) = \frac{\rho b}{N-1}$ and $\alpha(N) = \frac{\sigma b}{N-1}$.) As to the intercept $a^0(N)$, in some industries, $a^0(N)$ is increasing. In the restaurant industry, for example, the base demand level, for a given restaurant, under *given* price and service levels, i.e., the intercept in its demand function, often grows as additional restaurants are established in the area. The same may apply when an individual's utility of a service grows as the total number of users in the market increases (as is the case, for example, for internet access). Typically, $a^0(N)$ decreases in N, a phenomenon referred to as "business stealing" in the industrial organization literature.

When $a^0(N)$ is increasing in N, $w^{eq}(N)$ is decreasing in N, while $p^{eq}(N)$ is increasing in N. (Only the coefficient of the quadratic term in the cubic functions $C^*(w)$ depends on a^0. Thus, if $a^0(N)$ increases in N, note that the cubic functions $C^*(w)$ for $N = N_1$ and $N = N_2$, with $N_1 < N_2$, coincide when $w = 0$ while the former is pointwise larger than the latter. Thus, if the former ($N = N_1$) has a root on $[0, \overline{w}]$, the latter ($N = N_2$) has a smaller root. Either way, $w^{eq}(N_2) \leq w^{eq}(N_1)$, and it follows from (44) that $p^{eq}(N_2) \geq p^{eq}(N_1)$.) Thus, as more firms compete in the market, they offer better service but charge additionally for the service. It follows from (23), that each firm's demand volume increases with N as well.

When $a^0(N)$ is decreasing, the effects are reversed: $p^{eq}(N)$ is decreasing in N, while $w^{eq}(N)$ is increasing in N. Thus, as more firms compete in the market, the increased competition results in lower prices, but firms compensate by providing lower service as well. Again, by (23), decreasing equilibrium prices imply a lower demand volume per

Table 2: **Equilibria in the symmetric model,** $\eta = 1$

N	p^{eq}	w^{eq}	λ^{eq}	π^{eq}	Aggregate demand	Aggregate profit
2	2.83E+03	5.94E-03	2.83E+04	8.01E+07	5.66E+04	1.60E+08
3	1.91E+03	7.24E-03	1.90E+04	3.63E+07	5.71E+04	1.09E+08
4	1.44E+03	8.32E-03	1.44E+04	2.08E+07	5.77E+04	8.31E+07
5	1.17E+03	9.26E-03	1.16E+04	1.35E+07	5.82E+04	6.77E+07
10	6.10E+02	1.28E-02	6.08E+03	3.70E+06	6.08E+04	3.70E+07

Table 3: **Equilibria in the symmetric model,** $\eta = 2$

N	p^{eq}	w^{eq}	λ^{eq}	π^{eq}	Aggregate demand	Aggregate profit
2	1.44E+03	8.32E-03	1.44E+04	2.08E+07	2.88E+04	4.16E+07
3	6.72E+02	1.22E-02	6.70E+03	4.49E+06	2.01E+04	1.35E+07
4	4.02E+02	1.58E-02	4.00E+03	1.60E+06	1.60E+04	6.40E+06
5	2.77E+02	1.90E-02	2.75E+03	7.57E+05	1.38E+04	3.78E+06
10	1.10E+02	3.01E-02	1.08E+03	1.18E+05	1.08E+04	1.18E+06

firm. Whether *aggregate* sales decline or expand (a phenomenon referred to as "market expansion") depends on the rate at which a^0 declines with N. Consider the following numerical example:

Example 2: Consider the above symmetric model (with demand functions (43)). Let $a = b = 10, \rho = \sigma = 0.2$ and $c = \gamma = 1$. Finally, consider $a^0(N) = 1000 + 10^5/N^\eta$. $\eta > 0$. Table 2 below exhibits the equilibrium price, waiting time standard, sales volume and profits per firm , and the aggregate sales volume and profits, for $\eta = 1$ and $\eta = 2$. Since both $p^{eq}(N)$ and $\lambda^{eq}(N)$ decrease, gross profits, the first term in (7), decrease as well. While this may be somewhat offset by a reduction in the cost of the service based capacity - the second term in (7). we have, in all of the numerical explorations, observed that the *total* profit per firm reduces as well. Assuming this is the case, standard economic models would endogenize the number of firms N^*, as the largest value of N, for which the profit per firm exceeds a given critical level.

The numerical study in Allon and Federgruen (2004) investigates a variety of instances obtained from the base case in Example 1 (with $\epsilon = \eta = 0$) by varying one parameter at a time. The study focuses on 3 general managerial questions (I)-(III):

(I) Do firms necessarily benefit when some of the competitive choices can be made after other choices are revealed? (This is the case in the (SF) and (PF) models, compared to the (SC) model.)

(II) If a firm responds to a reduction of one of its cost parameters by offering a lower price as well as a better service level, will his competitors adjust their price and service level in the same direction?

(III) If customers become increasingly sensitive to the service level offered (as is the case in many service industries), will firms respond by offering higher prices and higher service levels and will they increasingly differentiate themselves along the service dimension?

Regarding (I), firms do *not* necessarily benefit when competing under (SF) or (PF) competition, compared to (SC). This phenomenon may arise even in settings, where under (SF), say, firms are guaranteed to offer higher prices (as well as service levels) along with higher demand volumes. In these settings, the uniformly larger demand volumes suggest that the *customers do* benefit from the sequential competition process. Likewise, hypothesis (II) is rejected by many examples. We have noted, in particular, that the reduction of one of a firm's cost parameters often allows this firm to simultaneously lower its price and improve its service level. Its competitors generally adjust their prices downward, as well, but sometimes need to compensate by compromising their service level. Finally, our admittedly limited numerical study confirms hypothesis (III).

Numerical explorations, reported in §2, have also shown that relatively small deviations from a full price instance may result in important changes in the equilibrium behavior of the industry. These instances, as well as many others reported in Allon and Federgruen (2004) illustrate how the impact of the "other attributes" may allow some firms to position themselves with *higher* prices and *lower* service levels than all its competitors, and nevertheless, maintain significant and sometimes even dominant market shares. We have also exhibited significant qualitative differences in the equilibrium behavior between the case in which the demand rates depend linearly on the service levels, and that in which the dependence is non-linear, reflecting decreasing marginal benefits to scale.

Future work should explore whether the above results continue to apply when service providers face more complex (than $M/M/1$) queuing systems, or when the demand functions follow one of the alternative specifications listed in §1.

Our results, to date, indicate that the characterizations of the equilibrium behavior in the (SC) model can be generalized for the Cobb-Douglas functions (CD). As for the attraction models (ATT), a Nash equilibrium can be guaranteed if the value functions are concave but not if they are merely log-concave, as in the MNL-type specification, where $\log v_i(p_i, \theta_i) = a_i\theta_i - b_i p_i$, with a_i and b_i positive constants. (Note that if the ratios a_i and b_i are identical, we recover the MNL specification in Cachon and Harker (2002)'s full price model. These authors showed that, even in this case, an equilibrium may fail to exist.) Leaving the comparative static results in §4.1 and §6.1 aside, the characterization of the equilibrium behavior in the price competition and the service competition model can, likewise, be extended for the non-linear demand structures (CD) and (ATT), among others. At the same time, characterization of the equilibrium behavior in the sequential competition models (SF) and (PF) is significantly harder to achieve. We have, in fact, encountered examples where it appears that a Nash equilibrium in the (SF) model, for example, fails to exist. More specifically, the tatônnement scheme for the (SF) model described in §4 can be used in order to identify Nash equilibria: whenever the scheme converges, its limit point is necessarily a Nash equilibrium. For an attraction model with

$$v_i = a_i^0 - b_i p_i + a_i \log(\theta_i), \tag{46}$$

the tatônnement scheme cycles between 8 distinct price-service level configurations irrespective of its starting point. (The starting point is generated randomly in the feasible region; the experiment has been repeated 1000 times.)

Example 3 Let $N = 3$, with demands given by the attraction model (46), where $a_1^0 = a_2^0 = 1800, a_1 = a_2 = 10, b_1 = b_2 = 15$, while $a_3^0 = 2700, a_3 = 15, b_3 = 20$ and $M = 1000$. The cost parameters are identical to those of examples 1 and 2. The tatônnement scheme cycles between the service level triples (1.63, 1.00, 2.89), (2.35, 1.00, 2.98), (5.05, 1.09, 3.34), (8.02, 1.18, 3.43), (8.20, 1.27, 3.52), (5.95, 1.36, 3.61), (6.31, 1.45, 3.7), (1.00, 1.00, 2.80).

We conclude that for some non-separable demand structures such as (ATT), *sequential* competition may not just result in a different equilibrium than *simultaneous* competition, but in fact may prevent the industry from settling at a stable equilibrium, in the first place. This is another manifestation of how knowledge about the industry's service levels, while determining price levels, may hinder rather than improve the industry's performance. One implication of the above observations is that when selecting a class of demand functions, e.g., (4), (CD) or (ATT), one should be guided not just by the tractability of estimation procedures and goodness-of-fit characteristics, but also by the implications for the industry's equilibrium behavior (and any prior knowledge thereof).

Future work should also consider generalizations of our model in which the firms' service processes are represented by more general queueing systems, or when customers are partitioned into several segments, each with its own price and service level for each firm. In addition, it would be desirable to integrate entry and exit decisions into the competition models.

References

Afeche, F. and H. Mendelson, (2004) "Pricing and Priority Auctions in Queueing Systems With Generalized Delay Cost Structure", forthcoming in *Management Science*

Allon, G. and A. Federgruen, (2002) "Competition in Service Industries", Working paper, Columbia University, NY, NY.

Anderson, S.P., A. de Palma and J.F. Thisse (1992) *Discrete Choice Theory and Product Differentiation*, The MIT Press, Cambridge, MA.

Bell, D., R. Keeny and J. Little, (1975). "A Market Share Theorem", *Journal of Marketing Research*, 12, 136-141.

Bernstein, F. and A. Federgruen, (2002) "Comparative Statics, Strategic Complements and Substitutes in Oligopolies", forthcoming in *Journal of Mathematical Economics.*

Blinder A. S., E. Canetti, D. Lebow (1998)*Asking About Prices: A New Approach to Understanding Price Stickiness*, Russell Sage Foundation, New York, NY.

Bowen B. and D. Headley (2001), "The Airline Quality Rating 2001", mimeo.

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

Carmon, Z. , J. G. Shanthikumar, and T. F. Carmon, (1994) "A Psychological Perspective on Service Segmentation: The Significance of Accounting for Consumers Perception of Waiting and Service, *Management Science*,41, 1806- 1815.

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

De Vany, A. and T. Savings, (1983) "The economics of quality", *Journal of Political Economy*, 91(6), 979-1000.

Fudenberg, D. and J. Tirole, (1984) "The Fat-Cat Effect, The Puppy-Dog Ploy, and the Lean and Hungry Look", *American Economic Review*, 74, 361-366.

Fudenberg, D. and J. Tirole, (1993) *Game Theory*, The MIT Press, Cambridge, Massachusetts.

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

Kahneman, D and A. Tversky, (1984) "Choices, Values and Frames", *American Psychologist*, 39, 341-350.

Kalai, E., M. Kamien and M. Rubinovitch, (1992) "Optimal service speeds in a competitive environment", *Management Science*, 38(8), 1154-1163.

Larson, R. C. (1987) "Perspectives on Queues: Social Justice and the Psychology of Queuing, *Operations Research*, 35 (6),895-904.

Leclerc, F., B. H. Schmitt, and L. Dube (1995) "Waiting Time and Decision Making: Is Time Like Money ?" *Journal of Consumer Research*, 22, 1 l0- 119.

Levhari, D. and I. Luski, (1978) "Duopoly pricing and waiting lines," *European Economic Review*, 11, 17-35.

Luski, I. (1976) "On partial equilibrium in a queueing system with two servers," *The Review of Economic Studies*, 43, 519-525.

Mazzeo, M. J. (2003). "Competition and Service Quality in the US Airline Industry", *Review of Industrial Organization*, 22 (4), 275-296

Milgrom, P. and J. Roberts, (1990) "Rationality, Learning and Equilibrium in Games with strategic Complementarities" *Econometrica*, 58 1255-1277.

So, K. C. (2000) "Price and time competition for service delivery" *Manufacturing Service* *and Operations Management* 2(4) 392-409.

Tirole, J. (1989) *The Theory of Industrial Organization*, The MIT Press, Cambridge, Massachusetts.

Topkis, D. M. (1998) *Supermodularity and Complementarity*, Princeton University Press. Princeton, NJ.

Vives, X. (2000) *Oligopoly Pricing: Old Ideas and New Tools*, The MIT Press, Cambridge, Massachusetts.