Pricing and Dimensioning Competing Large-Scale Service Providers

The literature on many-server approximations provides significant simplifications towards the optimal capacity sizing of large-scale monopolists but falls short of providing similar simplifications for a competitive setting in which each firm's decision is affected by its competitors' actions. In this paper, we introduce a framework that combines many-server heavy-traffic analysis with the notion of epsilon-Nash equilibrium and apply it to the study of equilibria in a market with multiple large-scale service providers that compete on both prices and response times. In an analogy to fluid and diffusion approximations for queueing systems, we introduce the notions of *fluid game* and *diffusion game*. The proposed framework allows us to provide first-order and second-order characterization results for the equilibria in these markets. We use our results to provide insights into the price and service-level choices in the market and, in particular, into the impact of the market scale on the interdependence between these two strategic decisions.

1. Introduction

In various service industries, an important attribute of the service experience is the delay experienced by customers who are waiting to be served. As a result, customers may consider both the price and the delay guarantees in choosing which provider to patronize. The purpose of this paper is to study the equilibria that emerge in markets in which large-scale service providers compete on both prices and service-levels. We focus on understanding the impact of the market size on the way in which different firms make their pricing and service-level choices.

Towards that end, we analyze a model of competition with multiple large-scale service providers in which the demand faced by each firm depends on the prices and the service levels offered by all firms in the market. Quantitatively, our goal is to characterize the capacity and pricing choices of the firms in the market. Qualitatively, we wish to understand how the strategic positioning of the firm depends on its own characteristics vis-à-vis those of its competitors.

To address these issues, we must first examine the firm's capacity decision. When service levels are measured through delays, a decision to improve the service level requires an investment in increased capacity. Hence, in positioning itself in the market, a firm must weigh the benefits of high service levels against the associated capacity costs. The benefits of improved service levels are not, however, independent of other competitors' actions and hence the task of determining the tradeoff between efficiency and service quality is a non-trivial one.

This tradeoff is a non-trivial one even for a monopolist. Indeed, when capacity is adjusted by determining the number of service representatives (rather than by adjusting the service rates) the problem of optimizing capacity costs vs. waiting-time-related costs is a complex optimization problem. While it can often be solved numerically, numerical solutions fail to provide any structural insights. An alternative to the exact numerical solutions is the use of approximations. Many-server approximations provide a simplified means to approach this problem; see e.g. Borst et al. (2004)–additional references are provided in §2. In this type of analysis, one considers a sequence of queueing systems with growing demand (and with capacity that grows accordingly to satisfy this demand). One then identifies solutions that are asymptotically optimal as the demand grows. The asymptotically optimal solution is nearly optimal for a given system provided that the demand it faces is large enough.

The literature on many-server approximations not only provides a tractable way to characterize nearly-optimal capacity and price decisions for monopolists; it also relates a firm's operational regime to the relative significance the firm ascribes to service levels as opposed to capacity costs; see the discussion of Borst et al. (2004) in §2. The firm's operational regime dictates how the firm should optimally respond to an increase in market size. Some firms should use their growth to increase their utilization (and thus their cost efficiency) without improving their service level. These firms are said to operate in the *Efficiency Driven* (ED) regime. Their emphasis on efficiency results in a situation in which (when the market is large) almost all customers experience some delay before being served. Some firms will sacrifice efficiency for quality. In response to an increase in market size, these firms will match an increase in utilization with yet a greater improvement in service levels. These operate in the *Quality Driven* (QD) regime. In this regime, almost all customers are immediately served. An intermediate regime is the Quality and Efficiency (QED) regime–also known as the Halfin-Whitt regime after the authors that first formalized it (see §2)–that corresponds to firms for which efficiency and quality are of similar importance. These firms will match the increase in efficiency with a comparable increase in the quality of service. In this regime, a non-trivial fraction (but not all) of the customers begins to receive service immediately, without any delay, but, at the same time, the efficiency is very high.

The regime-characterization results are proved in the literature for service providers that are monopolists in their respective markets. For a monopolist, the many-server approximations provide a tractable way to characterize its optimal capacity choices. The competitive setting is, however, more complex. Not only does the discrete nature of the capacity choice make the task of identifying equilibria and obtaining quantitative and qualitative results more arduous, the task is further complicated by the fact that the demand the firm experiences is not fixed, nor does it depend solely on the firm's own pricing and service-level choice. Rather, the demand depends on the choices made by all firms in the market. It seems plausible, however, that many-server approximations can be embedded within a game theoretic analysis to characterize equilibria in these markets. We pursue this direction by constructing a formal framework that draws on many-server approximations, as developed for monopolists, and by applying it to the study of equilibria in competitive markets.

Two fundamental questions are central to the study of equilibria in competitive markets: (a) *existence:* do Nash equilibria exist in the market?, and (b) *characterization:* given some sort of existence, is it possible to characterize the set of equilibria in order to obtain qualitative insights into the market outcomes? Starting with existence, we note that the concept of Nash equilibrium may be too restrictive for describing service-market behavior. It is known that Nash equilibria need not exist even under the most common demand functions, such as Multinomial Logit, and the simplest supply systems, such as the $M/M/1$ queue (see e.g. Cachon and Harker (2002)). This non-existence is often driven by economies of scale but is further exacerbated by the lumpy nature of the capacity in settings where the capacity choices are made in a discrete manner, by adjusting the number of service representatives. Non-existence of Nash equilibria does not rule out the possibility to say something meaningful about the market outcomes. It is desirable in these cases to find a less stringent framework that will allow for some characterization of the market outcomes.

The mathematical framework we propose is designed to address two concerns: (a) in terms of existence, we want to overcome the restrictive nature of the Nash equilibrium in addressing relatively general demand functions as well as supply facilities that are more general than the $M/M/1$ queue, and (b) in terms of characterization, we want to handle the complex nature of the service system by combining approximations for the queueing dynamics with a game theoretic framework.

Our framework stands on three pillars: (i) ϵ-Nash (or approximate) equilibria, (ii) many-server approximations, and (iii) market replication. The introduction of approximate equilibrium is aimed, initially, to overcome the non-existence of Nash equilibria. Its eventual benefits, however, go beyond this initial objective when combined with market replication and many-server approximations. We examine the behavior of equilibria, not on a single market, but rather on a sequence of markets with increasing aggregate demand – these are referred to as replicated markets. We emphasize that, 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.

Our framework can be thought of as a formalization of the use of fluid and diffusion models of queueing systems in a competitive setting. In the optimization of queueing systems, the original system is often replaced by a deterministic approximation – *a fluid model* – whose analysis sheds light on first-order properties of the underlying queueing system–such as its stochastic stability. In a second step, the original queueing system is replaced by a (more refined) stochastic model which is often referred to as a *diffusion model* of the queueing system. The latter is often more tractable than the original queueing system and can be used to identify properties that are asymptotically correct for the original queueing system. In particular, the diffusion model can be used to construct nearly optimal solutions for optimization problems that are intractable for the original queueing system.

Analogously, our framework constructs approximate games for the game played among the service providers. We first introduce a *fluid game* that is obtained from the original game by disregarding the stochastic nature of congestion. Building on the analysis of the fluid game, we then introduce a more refined *diffusion game*. This game is obtained from the original one by replacing each of the service providers with its many-server diffusion approximation. We then relate the equilibria of this new game with the outcomes of the original market. As in many-server approximations, the idea is to show that the equilibria of the diffusion game are, in a sense, asymptotically correct for the original game.

The notion of ϵ-Nash equilibrium plays a key role in rigorously establishing these approximations. The approximate equilibrium concept provides a formal way to construct *envelopes* for the profits of the firms in the market. While a Nash equilibrium might not exist and the market might oscillate, the ϵ-Nash identifies a region within which the *profits* of all the firms in the market must reside. The ultimate goal of this paper is, however, to understand market positioning in terms of the *actions* of the different firms, i.e, the prices and service levels that the firms choose. The challenge is, then, to use the envelopes on the profit functions to construct corresponding envelopes – in the action space – around the approximate price and service-level choices. To our knowledge, there are no general results that, given an ϵ-Nash equilibrium, identify the maximum that the firms can deviate in their *actions* without causing a deviation in the *profits* that would compromise the approximate equilibrium. Such results, that characterize the maximal oscillations of the prices and service-levels around some point are thus unique, and are obtained through the framework that we develop by employing the concepts of replicated markets in conjunction with heavy-traffic queueing theory.

Having constructed the analysis framework, we use it to provide an analytical characterization of the approximate equilibria in the market with multiple service providers. The characterization is then used to obtain some insights into the market outcomes. Our insights are concerned with the relationship between the price and service-level choices and, in particular, between the functions in the firm that make these choices – marketing and operations.

We identify *a one-sided decoupling* phenomenon by which the firms can be fairly close to optimality by allowing the price-setting function to "lead," and the operations function to "follow." The approximate equilibria–in both the fluid and diffusion level–exhibit a sequential structure: one can first pretend that the customers in the market are entirely insensitive to service levels and solve a simple price competition game. The real price and service level choices, in the approximate equilibrium, are then a function of this "naive" price vector but are, otherwise, independent of each other. This independence allows the marketing function to set the "naive" price vector as an initial estimate for the optimal price and leave for the operations function the task of setting the service levels and adjusting the prices that are eventually offered to the customers.

The analysis of the diffusion game provides a refined understanding of the operational regime of a firm and the implication of this regime on the firm's price choices. We show that both the Quality and Efficiency Driven (QED) and the Efficiency Driven (ED) regimes can emerge in equilibrium, thus, we appear to be the first to show how these different regimes emerge in a competitive market and, in particular, how different demand structures lead to the different regimes. We show that, while the actual choices of service level and price depend on the characteristics of all firms in the market, the operational regime of a firm is determined solely by its own intrinsic properties. Consequently, when different firms have different sensitivities, they may operate in different operational regimes, and thus position themselves differently in the face of increased market size.

We also find that the operational regime of a firm determines the degrees of freedom it has in pricing. We show that, compared with firms that operate in the ED regime, firms operating in the QED regime have greater freedom in choosing the prices they charge. Their freedom is reflected by the fact that they have a larger set from which they can choose their prices with hardly any compromise to their profits. Thus, firms in the QED regime can keep the one-sided decoupling in the sense that the marketing function can pay less attention to the operational side in determining the prices. Firms operating in the ED regime need to pay greater attention to their price choices. For these firms, the decoupling is weaker and a feedback mechanism is required between the manner in which the firm operates (i.e. the operational regime), and its pricing.

2. Literature review

Our work builds on two streams of literature: (a) game theory and its application to competition analysis, and (b) queueing theory and its application to the study of large-scale service systems. These two streams are not disjoint, and some recent work lies at the intersection of the two.

The literature on competition in service industries dates back to the late 1970s; see e.g. Levhari and Luski (1978). While it initially focused on a single attribute – price *or* service level (or a simple aggregation of the two), more recent work treats the prices and waiting-time standards as fully independent attributes. We follow Allon and Federgruen (2007) in considering a model with *differentiated* services, (i.e., a model in which other service attributes matter along with the full price) and in treating delay and price as independent attributes. We refer the reader to Allon and Federgruen (2007) for a systematic discussion of existing results in this context and to Hassin and Haviv (2003) for a general survey of queueing models with competition.

Allon and Federgruen (2007) and others focus on providing full analytical characterization of the Nash equilibria that arise in a market in which the market size is fixed. In contrast, we focus on understanding the impact of the market scale on the prices, service levels, and interdependencies between the two. Furthermore, our framework significantly expands the family of models that can be studied. This expansion is in two directions: (i) First, most of the literature on competition in services models the supply side via $M/G/1$ queues which implies, in turn, that capacity choices are made continuously by adjusting the service rates. We, in contrast, allow the service provider to adjust its capacity by increasing or decreasing the (integer-valued) number of service representatives, giving rise to an M/M/N queue, where N is a decision made by the firm. This is a common method of capacity management of service providers and one that renders Nash equilibrium intractable for characterization.(ii) Second, in terms of the demand model, our framework allows for significant generality in modeling the customers' sensitivity to service levels and prices.

From the game theoretic perspective, the notion of ϵ-Nash equilibria that we use has been used extensively in the economics literature. For the basic definition we rely on Tijs (1981). Dixon (1987) uses the idea of market replication in the context of price competition. While our form of replication is different, our analysis is inspired by his concept. Previous work in game theory has focused on four types of sequences of games: (i) sequences of games in which the action space is getting increasingly finer, and while each game has discrete action space, the limiting game has continuous action space (see e.g. Whitt (1980)), (ii) sequences of games in which the number of agents grows (Lu et al. (2007)), (iii) sequences of discrete time games in which the time between periods shrinks to zero along the sequence, and (iv) a sequence of replicated markets with growing market size (wee Dixon (1987)). We use the fourth framework.

The application of ϵ-Nash in the operations literature is rare. Lu et al. (2007) use this concept in a setting where Nash equilibrium does exist but the ϵ-Nash equilibrium concept still helps in characterizing the equilibrium in the game when the number of players is large and approaches a continuum. Dasci (2003) uses this concept in the context of ϵ-subgame-perfect equilibrium. We appear to be the first to combine the concepts of ϵ-Nash, market replication and heavy-traffic in the context of operational settings. This combination allows us to discuss both stability and trends in markets of competing service providers.

With respect to the relevant queueing literature, our work builds on the literature about many-server approximations of monopolists, starting with the seminal work of Halfin and Whitt (1981). While many-server approximations existed before, the result of Halfin and Whitt (1981) made such approximations relevant for various applications, such as call-center operations (see the survey papers by Gans et al. (2003) and Akşin et al. (2007)) and, more recently, health-care operations; see e.g. Jennings and de Véricourt (2008), Mandelbaum and Yom-Tov (2009).

Halfin and Whitt (1981) consider a sequence of $M/M/N$ queues and show that, as the demand rate Λ grows, the probability of delay $P\{W > 0\}$ converges to a number strictly between 0 and 1 if and only if the number of agents grows with Λ according to a *square-root safety staffing rule*, i.e, if and only if

$$N = R + \beta\sqrt{R} + o(\sqrt{R}), \tag{1}$$

where $R := \Lambda/\mu$ is the offered load and β is a strictly positive constant. In particular, a service provider that uses the square-root safety-staffing rule to determine his capacity will utilize his servers very efficiently and, at the same time, have a non-trivial fraction of its customers enter service immediately upon their arrival. This combination of high efficiency and high service level provides the justification for the name *Quality and Efficiency Driven* (QED) regime.

While Halfin and Whitt (1981) identified this regime, Borst et al. (2004) placed many-server approximations within a broad economic framework that considers the problem of minimizing capacity and waiting time costs. They show how the QED regime emerges as the optimal economical choice in some cases but also identify conditions under which other regimes, namely the Quality Driven (QD) and Efficiency Driven (ED) regime, emerge as the optimal choices. A key idea in the framework developed in Borst et al. (2004) is to replace the original optimization problem which involves the integer-valued number of servers by a tractable continuous and convex optimization problem. Using similar ideas, we will construct a continuous and tractable game–the *diffusion* *game*–that will serve as an approximation for the original, relatively intractable, one. Recently, Kumar and Randhawa (2008) extended the work of Borst et al. (2004) to a setting in which the customers are price-and-delay sensitive and consequently the demand is not fixed. Their work shows how different operational regimes emerge depending on the convexity (or concavity) of the delay-cost function. Similar dependencies will also emerge within the competitive setting that we study in this paper.

Additional examples of work that provides staffing and pricing recommendations for large-scale monopolists facing delay-and-price sensitive customers are the papers by Armony and Maglaras (2004), Whitt (2003) and Maglaras and Zeevi (2003, 2005).

All the work mentioned above considers a monopolist with a demand rate that may depend only on the congestion experienced and the price charged by this single player. Our paper appears to be the first to show how the different operational regimes emerge in a competitive setting with multiple players and to identify the dependencies between the operational regimes and the price choices of the various players in the market.

3. The model

We consider a market with a set $\mathcal{I} = \{1, \ldots, I\}$ of competing service firms, each operating as an $M/M/N$ facility and serving arriving customers in a First Come First Served (FCFS) manner. Firm i positions itself in the market by selecting a price p_i and a delay guarantee T_i. We restrict our attention to service-level guarantees that are given in terms of the customers' delay rather than their whole sojourn time in the system. Having chosen the delay target T_i, the service provider guarantees that the following Service Level (SL) constraint will be satisfied:

$$\mathbb{P}\{W_i > T_i\} \leq \phi, \tag{2}$$

where W_i is the steady-state delay with the i^{th} provider and $0 < \phi < 1$ is the satisfaction probability. This form of SL constraint is consistent with the industry practice that commonly uses $\phi = 0.2$ (corresponding to 80% of the service requests being answered within target; see e.g. Anton (2001)).[1] In this paper we study a model of competition where both the prices and the service levels are set simultaneously. We can show that our results continue to hold if the strategies are chosen sequentially (price first or service-level first).

[1]Our results are easily extended to the case where ϕ is allowed to vary between different firms.

Service rates are assumed to be fixed and equal to μ_i for firm i, and the capacities are adjusted through the choice of the number of agents (or service representatives), denoted by the integer-valued decision variable N_i. We assume that there is an upper bound $\bar{T} > 0$ on the acceptable service levels. For example, in call centers, it is clear that waiting time of more than a day is unacceptable. Firms choose $T_i \in [0, \bar{T}]$ and need to adjust their capacity, N_i, so as to guarantee that the SL constraint is satisfied for the chosen target. We let $\Theta := \times_{i=1}^{I}[0, \bar{T}]$.

Given the target T_i and the demand rate λ_i, the required capacity for firm i is given by

$$N_i = \min\left\{N \in \mathbb{Z}_+ : P\{W(\lambda_i, \mu_i, N) > T_i\} \le \phi\right\},$$

where $W(\lambda_i, \mu_i, N)$ is the steady-state delay in an $M/M/N$ queue with arrival rate λ_i, service rate μ_i and N servers.[2] We write

$$N_i = R_i + \hat{e}_i(\lambda_i, T_i), \tag{3}$$

where $R_i := \lambda_i/\mu_i$ is the offered load given the demand λ_i faced by firm i, and $\hat{e}_i(\lambda_i, T_i) := N_i - R_i$ is the excess capacity required to satisfy the service-level target. Naturally, we define $\hat{e}_i(\lambda_i, T_i) = 0$ whenever $\lambda_i = 0$ but we note that $\hat{e}_i(\cdot,\cdot)$ must be positive whenever $\lambda_i > 0$ to guarantee stability. The two terms in (3) represent the two components of the required capacity: the offered-load is the *volume-based capacity*, namely; it is the base capacity ensuring that the service process is stable. The second component ensures that the desired service levels are achieved and is referred to as the *service-based capacity*.

Firm i incurs a cost c_i per customer served and a cost γ_i per agent, per unit of time. This corresponds to the cost of capacity being linear in the number of agents.[3] The price p_i is chosen from a compact interval $[p_i^{min}, p_i^{max}]$, $i \in \mathcal{I}$. As each firm will select a price p_i which results in a non-negative gross profit margin $p_i - c_i - \gamma_i/\mu_i$, we assume, without loss of generality, that

$$p_i^{min} = c_i + \frac{\gamma_i}{\mu_i}\,, \quad i \in \mathcal{I}. \tag{4}$$

The upper bound, p_i^{max}, is allowed to obtain any value in $[p_i^{min}, \infty)$. We set $\mathcal{P}_i := [p_i^{\min}, p_i^{max}]$ and $\mathcal{P} := \times_{i=1}^{I}\mathcal{P}_i$. In full generality, the demand rates are specified as general functions of *all* prices and delay guarantees, i.e., $\lambda_i \equiv \lambda_i(p, T)$ where $p = (p_1, \ldots, p_I)$ and $T = (T_1, \ldots, T_I)$.

Assumption 3.1 (**regularity assumptions on the demand functions for differentiated services**) *For each $i \in \mathcal{I}$, the function $\lambda_i(\cdot,\cdot) : \mathcal{P} \times \Theta \mapsto \mathbb{R}_+$ is strictly positive, continuous and differentiable in all arguments and strictly decreasing in p_i and T_i.*

[2] N_i can be calculated by iteratively using the Erlang-C formula. Freeware calculators can be found, for example, at http://iew3.technion.ac.il/ serveng/ 4CallCenters/ Downloads.htm or http://www.cs.vu.nl/ koole/ ccmath/ ErlangC.

[3] See §8 for a discussion of more general capacity-cost models.

Firm-i's long-run-average profit Π_i, as a function of the prices and service levels in the market, is then given by

$$\Pi_i(p,T) = \lambda_i(p,T)(p_i - c_i) - \gamma_i N_i,$$

which, using (3), is re-written as follows:

$$\Pi_i(p,T) = \lambda_i(p,T)\left(p_i - c_i - \frac{\gamma_i}{\mu_i}\right) - \gamma_i \hat{e}_i(\lambda_i, T_i)\ . \tag{5}$$

The assumption of large-scale service systems is introduced by considering a family of markets indexed by a market-scale multiplier $\Lambda \geq 0$ so that the demand grows with the market-scale multiplier in a natural way. Specifically, we let

$$\Lambda_i(p,T) := \Lambda \cdot \lambda_i(p,T), \tag{6}$$

be the demand facing firm i in the Λ^{th} market. The profit functions in the Λ^{th} market are then given by

$$\Pi_i^{\Lambda}(p,T) = \Lambda_i(p,T)\left(p_i - c_i - \frac{\gamma_i}{\mu_i}\right) - \gamma_i \hat{e}_i(\Lambda_i, T_i)\ ,\ i \in \mathcal{I}. \tag{7}$$

For future reference we make the following formal definition

Definition 3.1 (**the market game**) The Λ^{th} market game is the I-player game with profit functions $\{\Pi_i^{\Lambda}(\cdot,\cdot)$, $i \in \mathcal{I}\}$ and strategy space $\mathcal{P} \times \Theta$.

As is the case in heavy-traffic analysis, the key idea of our market procedure is to embed the real market (with fixed market size) into a sequence of markets with growing demand. If one is able to get meaningful results for the sequence of markets, these can be applied to a market with fixed size as long as the size is large enough. Looking at the sequence of markets, we are interested in understanding how the stability of the market and the market outcomes change with the increase in market size. Following conventional notation we let

$$(p,T)_{-i} = ((p_1,T_1),\ldots,(p_{i-1},T_{i-1}),(p_{i+1},T_{i+1}),\ldots,(p_I,T_I)).$$

We denote by $T_i^{*,\Lambda}(p,T)$ and $p_i^{*,\Lambda}(p,T)$, respectively, the delay and price components of firm's i best response to $(p,T)_{-i}$ in the Λ^{th} market game. The existence of a best response for any actions $(p,T)_{-i}$ follows from the continuity of the demand functions and the compactness of the strategy space. When the best response is not unique, we arbitrarily choose one best response. The way this best response is chosen will be immaterial for our results.

As discussed in the introduction, the market game is intractable for direct Nash equilibria analysis. This is a consequence of the complexity of the expressions for the service-based capacity, the discreteness of this capacity and the concavity of the capacity-cost function.[4] Instead, we take an indirect approach that exploits the benefits of large-scale asymptotic analysis within an ϵ-Nash-equilibrium framework.

3.1 ϵ-Nash equilibria

The notion of ϵ-Nash equilibria is adopted from Tijs (1981). Rather than defining it in general terms, we provide the definition as it applies to our setting. To this end, given $(p, T) \in \mathcal{P} \times \Theta$ and $(\tilde{p}_i, \tilde{T}_i) \in \mathcal{P}_i \times [0, T]$, we let

$$(\tilde{p}_i, \tilde{T}_i) \uparrow (p, T)_{-i} = ((p_1, T_1), \ldots, (p_{i-1}, T_{i-1}), (\tilde{p}_i, \tilde{T}_i), (p_{i-1}, T_{i-1}), \ldots, (p_I, T_I)).$$

Definition 3.2 (ϵ**-Nash equilibrium for the** Λ^{th} **market game**) *Fix* $\Lambda \geq 0$. *Let* $\epsilon = (\epsilon_1, \ldots, \epsilon_I)$ *be a positive vector. We say that* $x \in \mathcal{P} \times \Theta$*, is an* ϵ*-Nash equilibrium of the* Λ^{th} *market game if, for each* $i \in \mathcal{I}$ *and any* $\tilde{x}_i \in \mathcal{P}_i \times [0, \bar{T}]$,

$$\Pi_i^{\Lambda}(\tilde{x}_i \uparrow x_{-i}) \leq \Pi_i^{\Lambda}(x) + \epsilon_i \ .$$

Nash equilibrium is a special case of ϵ-Nash in which $\epsilon = 0$. The generalization from Nash to ϵ-Nash allows us to construct an "envelope" around the market outcomes and thus obtain key insights about the market behavior even in cases in which Nash equilibria do not exist. The ability to construct such "envelopes" is useful also in cases in which Nash equilibria do exist but are difficult to characterize. In these cases, if ϵ is small enough, the characterization of the ϵ-Nash equilibria can shed light on the Nash equilibrium. We will be formally constructing such "envelopes" as well as analyzing the gaps between the ϵ-Nash and Nash equilibria whenever the latter exist.

Notational conventions and organization of the paper: for two sequences of positive vectors $\{a^{\Lambda},\ \Lambda \geq 0\}$ and $\{b^{\Lambda},\ \Lambda \geq 0\}$ with elements in $\mathbb{R}^d_+$ we say that $a^{\Lambda} = O(b^{\Lambda})$ if $\limsup_{\Lambda} a_i^{\Lambda}/b_i^{\Lambda} < \infty$ for $i = 1, \ldots, d$. We say that $a^{\Lambda} = o(b^{\Lambda})$ if $\limsup_{\Lambda} a_i^{\Lambda}/b_i^{\Lambda} = 0$ for $i = 1, \ldots, d$. Finally, we say that $a^{\Lambda} \sim b^{\Lambda}$ if $a^{\Lambda} = O(b^{\Lambda})$ but $a^{\Lambda} \neq o(b^{\Lambda})$. For a vector $x \in \mathbb{R}^d$, we let $\|x\| = \sum_{k=1}^{d} |x_k|$. When

[4] It is possible to construct continuous versions of the service-based capacity–see e.g. §4 of Borst et al. (2004). This, however, this would still leave the market game intractable for exact analysis

applied to a vector $x \in \mathbb{R}^d$, the absolute value operation should be interpreted componentwise, i.e., $|x| = (|x_1|, \ldots, |x_d|)$. Similarly, the square-root operator should be interpreted componentwise, i.e, for $x \in \mathbb{R}^d_+$, $\sqrt{x} = (\sqrt{x_1}, \ldots, \sqrt{x_d})$. The notation "$\to$" stands for convergence as $\Lambda \to \infty$ unless explicitly stated otherwise. We will often use 0 to represent the 0 vector in $\mathbb{R}^d$ and the dimension of the vector will always be clear from the context.

The rest of the paper is organized as follows: §4 is concerned with regime characterization. In §5 we introduce and characterize the fluid game and then discuss its implications. In §6 we turn to the diffusion game which is concerned with a refined understanding of the firms' choices. In §7 we provide a detailed illustration of our results using a linear demand model. Conclusions and directions for future research are discussed in §8. The appendix contains additional numerical examples as well as generalizations to some of the results in §5 and 6.

Our approach in presenting the results is to state them formally within the paper, accompanied by various examples for illustration. Most of the detailed proofs are relegated to the e-companion.

4. Regime characterization

In this section we discuss the optimal operational regimes of the firms in the market and relate this regime choice to the underlying demand models. The firm's operational regime characterizes how the firm behaves in the face of an increased market size. The literature identifies three possible regimes: (i) firms are said to operate in the *Efficiency Driven* (ED) regime if they use demand-growth to increase their utilization (and thus their cost efficiency) without improving their service level; (ii) firms are said to operate in the *Quality Driven* (QD) regime if, in response to an increase in market size, they match the increase in utilization with yet a greater improvement in service levels, and finally (iii) firms are said to operate in the *Quality and Efficiency Driven* (QED) regime if they match the increase in efficiency with a comparable increase in the quality of service. A firm's operational regime is the outcome of the firm trading off its capacity cost and the service level it provides. For a monopolist, this tradeoff is solely a function of the firm's own scale economies. In an oligopolistic setting, however, the value of a service level for a given firm depends on its competitors' decisions, thus making the tradeoff more subtle.

The outcome of this initial analysis will be a mapping from firm i's demand structure to a quantifier r_i^Λ. This quantifier characterizes the order of magnitude of the optimal service-level choice for firm i. Some firms will have $r_i^\Lambda = 1/\sqrt{\Lambda}$ and we will show that, for these firms, it is optimal to use a service-based capacity that is of the order of $\sqrt{\Lambda}$. Consequently, these firms will operate (in equilibrium) in the QED regime. Other firms will have r_i^Λ which is significantly larger than $1/\sqrt{\Lambda}$. These firms will optimally use a service-based capacity of a magnitude $o(\sqrt{\Lambda})$ and will operate in the ED regime.

To motivate our results of regime characterization, note that, given a price vector p^Λ and the service-level choices T_{-i}^Λ of its competitors, firm i's best service-level choice is given by

$$T_i^\Lambda \in \operatorname*{argmax}_{x\in[0,\bar{T}]} \Pi_i^\Lambda(p^\Lambda, T_{-i}^\Lambda, x) = \operatorname*{argmax}_{x\in[0,\bar{T}]} \Lambda_i(p^\Lambda, T_{-i}^\Lambda, x)\left(p_i^\Lambda - c_i - \frac{\gamma_i}{\mu_i}\right) - \gamma_i \hat{e}_i(\Lambda_i, x).$$

Here and henceforth we use (p, T_{-i}, T_i) to denote the vector $(p_i, T_i) \uparrow (p, T)_{-i}$. Equivalently,

$$T_i^\Lambda \in \operatorname*{argmax}_{x\in[0,\bar{T}]} \left[\Lambda_i(p^\Lambda, T_{-i}^\Lambda, x) - \Lambda_i(p^\Lambda, T_{-i}^\Lambda, 0)\right]\left(p_i^\Lambda - c_i - \frac{\gamma_i}{\mu_i}\right) - \gamma_i \hat{e}_i(\Lambda_i, x). \tag{8}$$

The order of magnitude of T_i^Λ is determined, then, by optimally balancing the loss of market share due to customer delays–which we informally refer to as the "delay cost"–and is given by $\Lambda_i(p, T_{-i}, x) - \Lambda_i(p, T_{-i}, 0)$ and the service-based capacity cost $\gamma_i \hat{e}_i(\Lambda_i, x)$[5]. Assumption 4.1 below provides us with some control of the delay cost and Lemma 4.1 provides estimates on the order of magnitude of the service-based capacity.

Lemma 4.1 *Fix a sequence* $\{(p^\Lambda, T^\Lambda),\ \Lambda \geq 0\}$ *such that* $(p^\Lambda, T^\Lambda) \in \mathcal{P} \times \Theta$ *for all* $\Lambda \geq 0$. *Then,*

$$\hat{e}_i(\Lambda_i, T_i^\Lambda) \sim \min\left\{\frac{1}{T_i^\Lambda}, \sqrt{\Lambda}\right\}.$$

Assumption 4.1 (**behavior around** $T = 0$) *For each* $i \in \mathcal{I}$ *there exists* $\alpha_i > 0$ *such that*

$$\limsup_{x\to 0} \sup_{p, T_{-i}} \frac{\lambda_i(p, T_{-i}, 0) - \lambda_i(p, T_{-i}, x)}{x^{\alpha_i}} < \infty, \text{ and } \liminf_{x\to 0} \inf_{p, T_{-i}} \frac{\lambda_i(p, T_{-i}, 0) - \lambda_i(p, T_{-i}, x)}{x^{\alpha_i}} > 0.$$

Note that, returning to (8) and using Assumption 4.1 and Lemma 4.1, we have (informally) that

$$T_i^\Lambda \sim \operatorname*{argmax}_{x\in[0,\bar{T}]} \left[-\Lambda x^{\alpha_i} - \min\left(\frac{1}{x}, \sqrt{\Lambda}\right)\right].$$

Hence, we should have that $T_i^\Lambda \sim r_i^\Lambda$ where

$$r_i^\Lambda := \operatorname*{argmin}_{x\in\left[\frac{1}{\sqrt{\Lambda}}, \bar{T}\right]} \Lambda x^{\alpha_i} + \frac{1}{x}.$$

[5]Here, the loss of market share parallels the role of the delay cost in the monopolist setting of Borst et al. (2004).

A simple calculation then yields

$$r_i^\Lambda = \max\left\{\frac{1}{\Lambda^{\frac{1}{1+\alpha_i}}}, \frac{1}{\sqrt{\Lambda}}\right\}, \tag{9}$$

so that $r_i^\Lambda = 1/\sqrt{\Lambda}$ for all $\alpha_i \leq 1$ and $r_i^\Lambda = \Lambda^{-\frac{1}{1+\alpha_i}}$ otherwise.

Assumption 4.1 requires that, in the vicinity of $T = 0$, the demand volume of a firm decreases proportionally to **some** power of its delay guarantee T_i. The power may be different for different firms. This assumption is satisfied by most known demand models, but may not be satisfied in general; see the discussion and examples at the end of this section.

The quantifier r_i^Λ, which depends on Assumption 4.1 through (9), plays an important role in determining the operational regime of a firm. As suggested by the informal discussion so far, r_i^Λ provides an order-of-magnitude estimate for the service-level choice of firm i and $1/r_i^\Lambda$ provides an estimate of the service-based capacity for that firm i. This is formally stated in the following theorem.

Theorem 4.2 (**regime characterization**) *Suppose that Assumption 4.1 holds. Let* $\{(p^\Lambda, T^\Lambda),\ \Lambda \geq 0\}$ *be a sequence such that* $(p^\Lambda, T^\Lambda) \in \mathcal{P} \times \Theta$ *for all* $\Lambda \geq 0$*. Then,*

$$T_i^{*,\Lambda}(p^\Lambda, T^\Lambda) \sim r_i^\Lambda,\ i \in \mathcal{I} : \alpha_i > 1, \tag{10}$$

$$T_i^{*,\Lambda}(p^\Lambda, T^\Lambda) = O(r_i^\Lambda),\ i \in \mathcal{I} : \alpha_i = 1, \tag{11}$$

and

$$T_i^{*,\Lambda}(p^\Lambda, T^\Lambda) = o(r_i^\Lambda),\ i \in \mathcal{I} : \alpha_i < 1. \tag{12}$$

Furthermore,

- **QED regime:** *if* $\frac{1}{r_i^\Lambda} \sim \sqrt{\Lambda}$*, then*

  $$N_i^\Lambda - \frac{\Lambda_i}{\mu_i} \sim \sqrt{\Lambda},$$

  and

  $$\limsup_{\Lambda\to\infty} P\{W_i^\Lambda > 0\} < 1, \quad \textit{and} \quad \liminf_{\Lambda\to\infty} P\{W_i^\Lambda > 0\} > 0.$$

- **ED regime:** *If* $\frac{1}{r_i^\Lambda} = o(\sqrt{\Lambda})$*, then*

  $$N_i^\Lambda - \frac{\Lambda_i}{\mu_i} = o(\sqrt{\Lambda}),\ \textit{and} \lim_{\Lambda\to\infty} P\{W_i^\Lambda > 0\} = 1.$$

  Here, N_i^Λ *is firm* i*'s capacity under a best response to* (p^Λ, T^Λ) *in the* Λ^{th} *market game and* W_i^Λ *is the steady-state delay at firm* i *under this best response.*

Interestingly, Theorem 4.2 implies that even if the market oscillates between different points in $\mathcal{P} \times \Theta$, the operational regime of a firm remains unchanged. Moreover, it shows that, while the actual choice of service level by a firm depends on the characteristics of all firms in the market, its operational regime–Efficiency Drive (ED), Quality Driven (QD) or Quality and Efficiency Driven (QED)– depends only on its own intrinsic properties as reflected in the quantifier r_i^Λ.

Remark 4.3 (**the Quality-Driven (QD) regime**) The QD regime, in which the probability of delay, $P\{W^\Lambda > 0\}$, approaches 0 as $\Lambda \to \infty$ does not emerge in our setting. This is a consequence of the structure of the service-level constraints that we use in our model. Specifically, when a firm's service level is defined via $P\{W_i > T_i\} \leq \phi$ for ϕ that is strictly positive and exogenously given, it can not do better than setting $T_i = 0$. In this case, Proposition 1 in Halfin and Whitt (1981) tells us that, in order to have $P\{W_i > 0\} \leq \phi$ for $\phi \in (0,1)$, it suffices to use the square-root-safety staffing rule and, in particular, to use a service-based capacity that is proportional to the square-root of the demand. Hence, it cannot be optimal for a firm to operate in the QD regime which requires the service-based capacity to be orders of magnitude greater than $\sqrt{\Lambda}$. The framework that we provide in this paper can, however, be applied to alternative settings in which the QD regime does emerge in competition. We expect, for example, that, if service levels are defined via guarantees of the form $E[W_i] \leq T_i$, the QD regime will emerge as a possible outcome. Indeed, under some demand models it may be optimal for some firms to guarantee an average delay T_i^Λ such that $T_i^\Lambda = o(1/\sqrt{\lambda})$. This, in turn, would imply that the corresponding firm is operating in the QD regime; see §9 of Borst et al. (2004); see §9 of Borst et al. (2004). ■

We conclude this section by pointing out some widely used demand models that satisfy Assumption 4.1. The multinomial logit and the Cobb-Douglas models are two such examples.

Example 4.1 (**the multinomial logit (ML) demand model**) Fix $i \in \mathcal{I}$ and assume that

$$\lambda_i(p,T) := \frac{e^{a_i(T_i) - b_i p_i}}{v_0 + \sum_j e^{a_j(T_j) - b_j p_j}}, \tag{13}$$

where $v_0 > 0$ is a constant and $a_i(T_i) = a_i - k_i(T_i)^{\alpha_i}$, for a_i, k_i and α_i being positive constants. Then, it can be easily verified that α_i in the definition of $a_i(T_i)$ plays the role of the exponent in Assumption 4.1. It is important to note that in a market in which the demand experienced by each firm is characterized by the Multinomial Logit model, some firms may be operating under either the ED regime and some in the QED regime, depending on the sensitivity of the attraction values of each firm to its own service-level. ■

Henceforth, whenever we mention the ML demand model we will be referring to the one in Example 4.1.

Example 4.2 (**demand models with Taylor expansion around** $T = 0$) A large family of models for which Assumption 4.1 is satisfied are those in which the function $\lambda_i(p, T)$ has a Taylor series expansion around $T_i = 0$. In these cases, expanding around $T_i = 0$, we can write

$$\lambda_i(p, T_{-i}, x) = \lambda_i(p, T_{-i}, 0) + \sum_{l=1}^{k} \frac{\partial^l}{\partial^l T_i} \lambda_i(p, T)\Big|_{T_i=0} \cdot x^l + o(x^k).$$

Assumption 4.1 is then satisfied with an exponent that corresponds to the first non-zero derivative with respect to T_i at the point $T = 0$ provided that the derivative is uniformly bounded away from 0. Formally, we will have $\alpha_i = k$ where k is such that

$$\frac{\partial^k}{\partial^k T_i} \lambda_i(p, T)\Big|_{T_i=0} \in (-a, -b), \quad \text{and} \quad \frac{\partial^l}{\partial^l T_i} \lambda_i(p, T)\Big|_{T_i=0} = 0, \; l < k, \tag{14}$$

for some $0 < a < b < \infty$ and for any vectors p and T_{-i}. Of course, Assumption 4.1 holds also in many examples in which such a Taylor expansion does not exist; see Example 4.1 above. In that example a Taylor expansion as in (14) need not exist when $a_i(T_i) = a_i - \kappa_i(T_i)^{\alpha_i}$ for a non-integer exponent α_i. ■

The following example illustrates a demand model for which all parameter values lead to the exponent $\alpha_i = 1$ and, consequently, to the QED regime.

Example 4.3 (**the Cobb-Douglas demand model**) Fix $i \in \mathcal{I}$ and assume that

$$\lambda_i(p, T) := \frac{v_i(p_i, T_i)}{v_0 + \sum_j v_j(p_j, T_j)} \; ,$$

with $v_i(p_i, T_i) := c_i \left(\frac{\bar{T}}{\bar{T} - T_i}\right)^{-a_i} p_i^{-b_i}$ for strictly positive constants a_i, b_i and c_i. Here, it is easy to verify that equation (14) holds with $k = 1$ so that we always have, $\alpha_i = 1$ in Assumption 4.1. Consequently, in a market in which all the firms face a Cobb-Douglas demand model, only the QED regime emerges as the optimal choice. ■.

Roadmap for rest of the paper: In sections 5 and 6 we introduce the *fluid game* and the *diffusion game*, respectively. The fluid game provides a first-order characterization of the market outcome. It will also serve an essential building block in the introduction and characterization of the, more refined, diffusion game. The fluid game and diffusion game will differ from each other, and from

Game	Strategy space	Payoff function
Market	$\mathcal{P} \times \Theta$	$\Pi_i^\Lambda(p,T)$ as in (7)
Fluid	$\mathcal{P}$	$\bar{\Pi}_i^P(p) := \lambda_i(p,0)\left(p_i - c_i - \frac{\gamma_i}{\mu_i}\right)$
Diffusion	$\mathcal{P} \times \Theta$	$\hat{\Pi}_i^\Lambda(p,T) := \lambda_i(p,T)\left(p_i - c_i - \frac{\gamma_i}{\mu_i}\right) - \gamma_i \Lambda f_i(T_i)$

Table 1: Three different games

the market game of Definition 3.1, in terms of the payoff functions and the strategy spaces as outlined in Table 4. The exact derivation of the payoff functions as well the definition of the function $f_i(\cdot)$ in the diffusion-game payoff will be specified explicitly in §5 and in §6. It is important to note that the strategy space of the fluid game is only the price domain $\mathcal{P}$ and no service-based capacity cost appears in the payoff function of this approximate game. It is also important that the service-based capacity cost for firm i in the diffusion game depends only on service-level choice of the specific firm. This stands in contrast to the market game in which the service-based capacity is $\hat{e}_i(\Lambda_i(p,T),T_i)$ and hence depends on the complete vector (p,T). For each of the above approximate games we will characterize the equilibria: a single equilibrium $(p^*,0)$ for the fluid game, and a sequence of equilibria $\{(p^\Lambda,T^\Lambda),\Lambda \geq 0\}$ for the sequence of diffusion games. We will show that, at different levels of precision, these equilibria approximate the original-market-game outcomes. The precision of the approximation will be a function of the quantifiers r_i^Λ identified earlier in this section. We will measure the quality of these approximations in two dimensions: in *payoff space* and in *action space*. In payoff space we will measure the maximum profitable deviation any firm can achieve *in terms of profits* if the market is initialized with the proposed approximate equilibrium (the fluid-game equilibrium or the diffusion-game equilibrium). We will identify sequences ϵ^Λ such that the maximum profitable deviation (in payoffs) is smaller than ϵ^Λ for the Λ^{th} market game.

We will then translate the payoff-space bounds to action-space bounds. Namely, starting the market in one of the approximate equilibria, we will characterize the maximum profitable deviation that a firm will deviate in its actions–service level and price. Accordingly, with the sequence of markets, we will associate a sequence of bounds on actions. We will identify conditions under which the service-level choice in the diffusion-game equilibrium is a precise prediction of the market outcomes up to an order of $o(r_i^\Lambda)$ for firm i.

To translate the bounds in payoff space to bounds in the action space $\mathcal{P} \times \Theta$, we will have to impose certain assumptions and conditions. Interestingly, by imposing conditions on the fluid game alone, we will be able to establish bounds in action space for both the fluid and the diffusion game. This is important as the fluid game is a game with relatively simple payoff functions and with the simple strategy space $\mathcal{P}$. For each of the assumptions and conditions that we impose we will provide concrete examples of commonly used demand models that satisfy these conditions.

Table 2 summarizes the results that will appear in the next two sections. The last column provides a list of the assumptions and conditions that will be required for each result. The conditions (C1)-(C5) will be formally introduced the first time they are used. The table should be used as a roadmap and a reference point in reading through the rest of the paper.

Game	Equilibrium	Quality of appr. Payoff Space	Quality of appr. Action Space	Assumptions & conditions	Theorems
Fluid	$(p^*, 0)$	$\epsilon_i^\Lambda = o(\Lambda)$	$o(1)$	Assumptions 3.1 and 5.1	5.3
Fluid	$(p^*, 0)$	$\epsilon_i^\Lambda = O(1/r_i^\Lambda)$	$O(r^\Lambda)$ for service $O(B^{-1}\delta^\Lambda)$ for price	Assumptions 3.1, 4.1 and 5.1 Conditions (C1)-(C3)	5.6, 5.9
Diffusion	(p^Λ, T^Λ)	$\epsilon_i^\Lambda = o(1/r_i^\Lambda)$	$o(r^\Lambda)$ for service $o(B^{-1}\zeta^\Lambda)$ for price	Assumptions 3.1, 4.1 and 5.1 Conditions (C1)-(C5)	6.3, 6.7

Table 2: Summary of results

The quantities δ_i^Λ, ζ^Λ in Table 2 are simple functions of the quantifier r^Λ and will be specified explicitly in the corresponding results. The matrix B is independent of Λ and will be introduced in the next section §5.

5. A fluid game

The *fluid game* that we introduce in this section is a simplification of the market game. In this simplified game only the first-order impact of the firms' actions is modeled. This is achieved by replacing the service facilities (the $M/M/N$ queues) by their fluid approximations. We will show that this game does indeed provide a first-order approximation to the original game in that the market prices will always lie within some small neighborhood of the fluid-game equilibrium and the service levels will be close, in a sense, to $T = 0$.

5.1 Definition and characterization

Definition 5.1 (**the fluid game**) *The fluid game is the I-player game with profit functions*

$$\bar{\Pi}_i^P(\cdot) := \lambda_i(p, 0) \cdot \left(p_i - c_i - \frac{\gamma_i}{\mu_i} \right), \; i \in \mathcal{I},$$

and strategy space $\mathcal{P}$.

Note that the fluid game has the original (unscaled) demand functions $\{\lambda_i(\cdot),\ i \in \mathcal{I}\}$. In the fluid game, the players compete only on prices, i.e, this is a pure-price competition game. Furthermore, the strategy space, $\mathcal{P}_i$, of each player is a compact subset of $\mathbb{R}_+$ so that there exist numerous sufficient conditions for the existence and uniqueness of equilibria. For existence, it suffices to have that $\bar{\Pi}_i^P(\cdot)$ is continuous and quasi-concave with respect to p_i (see §2.3 of Cachon and Netessine (2004)). This sufficient condition is guaranteed, for example, for attraction models such as the Multinomial Logit demand model or the Cobb Douglas demand model; see Examples 4.1 and 4.3. We will assume that there is a unique equilibrium (and discuss later some concrete examples in which this assumption indeed holds). We formally state these requirements in the following assumption.

Assumption 5.1 (**existence and uniqueness of equilibrium for the fluid game**) *The fluid game has a unique Nash equilibrium* $p^* := (p_1^*, \ldots, p_I^*)$.

For the rest of the paper, whenever Assumption 5.1 holds, we will use the notation p^* when referring to the unique equilibrium of the fluid game.

5.2 The quality of the approximation

Theorems 5.2 and 5.3 below show that the fluid game serves as a first-order approximation for the original market game.

Theorem 5.2 (**existence of approximate equilibria**) *Suppose that Assumptions 3.1 and 5.1 hold and let* $\{\epsilon^\Lambda,\ \Lambda \geq 0\}$ *be a sequence of vectors in* $\mathbb{R}_+^I$ *that satisfies*

$$\frac{\epsilon_i^\Lambda}{\Lambda} \to 0 \text{ and } \epsilon_i^\Lambda \to \infty,\ i \in \mathcal{I}. \tag{15}$$

Then, there exists a sequence $\{T^\Lambda,\ \Lambda \geq 0\}$ *such that* $T_i^\Lambda \to 0$ *for all* $i \in \mathcal{I}$ *and for each* Λ*, the vector*

$$(p^*, T^\Lambda) = ((p_1^*, T_1^\Lambda), \ldots, (p_I^*, T_I^\Lambda)),$$

is an ϵ^Λ*-Nash equilibrium for the* Λ^{th} *market game. Moreover,* T^Λ *can be chosen so that* $T_1^\Lambda = T_2^\Lambda = \ldots = T_I^\Lambda$.

While Theorem 5.2 shows that an approximate equilibrium exists, the following theorem shows that all approximate equilibria must be contained within a small neighborhood of $(p^*, 0)$.

Theorem 5.3 (**first-order characterization**) *Suppose that Assumptions 3.1 and 5.1 hold and let* $\{(p^\Lambda, T^\Lambda), \Lambda \geq 0\}$ *and* $\{\epsilon^\Lambda, \Lambda \geq 0\}$ *be such that* (p^Λ, T^Λ) *is an* ϵ^Λ*-Nash equilibrium for the* Λ^{th} *market game and such that* (15) *holds. Then, as* $\Lambda \to \infty$*,*

$$T_i^\Lambda \to 0,\ i \in \mathcal{I}, \tag{16}$$

and

$$p_i^\Lambda \to p_i^*,\ i \in \mathcal{I}. \tag{17}$$

Theorems 5.2 and 5.3 are driven by economies of scale. Using Lemma 4.1, we see that the service-based capacity $\hat{e}_i(\cdot,\cdot)$ grows at a lower rate than the volume-based capacity–even for small delay guarantees. In particular, for any sequence $\{(p^\Lambda, T^\Lambda),\ \Lambda \geq 0\}$ with $(p^\Lambda, T^\Lambda) \in \mathcal{P} \times \Theta$, $\hat{e}_i(\Lambda_i, T_i^\Lambda) = o(\Lambda_i)$. Consequently, the profit functions satisfy the following property:

$$\Pi_i^\Lambda(p^\Lambda, T^\Lambda) = \Lambda_i(p^\Lambda, T^\Lambda)\left(p_i - c_i - \frac{\gamma_i}{\mu_i}\right) + o(\Lambda_i).$$

Due to the relatively low cost of the service-based capacity one expects the firms to choose to provide relatively high service levels (corresponding to small values of T_i). Accordingly, we expect that a game with profit functions

$$\Pi_i^{\Lambda,P}(p) := \Lambda_i(p,0)\cdot\left(p_i - c_i - \frac{\gamma_i}{\mu_i}\right) = \Lambda\cdot\lambda_i(p,0)\cdot\left(p_i - c_i - \frac{\gamma_i}{\mu_i}\right),\ i \in \mathcal{I}, \tag{18}$$

and strategy space $\mathcal{P}$ will provide a first-order approximation for the Λ^{th} market game. Division by the common scalar Λ yields the fluid game in Definition 5.1.

Combined, Theorems 5.2 and 5.3 show that the market prices and service-levels must reside within an increasingly small envelope around $(p^*, 0)$. In particular, if a Nash equilibrium (p^Λ, T^Λ) exists for the Λ^{th} market game for each Λ, then the sequence of these Nash equilibria must converge to $(p^*, 0)$. We note that Theorems 5.2 and 5.3 do not characterize the convergence rate of T_i^Λ to 0 and of p_i^Λ to p_i^*, nor do they relate this convergence rate to the bounds ϵ^Λ on the profit functions. Relating these convergence rates is our ultimate goal in this section. Before moving towards that goal, we discuss some practical implications of the fluid game.

Remark 5.4 (**interpreting** ϵ^Λ **as the level of sub-optimality**) One may interpret ϵ as the level of sub-optimality for a firm if it chooses to price according to the ϵ-Nash-equilibrium price. To illustrate this point, consider the special case in which the market has a single firm–a monopolist.

The implication of Theorem 5.2 for this special case is that the monopolist cannot increase its profit by more than ϵ^{Λ} by deviating, i.e, that

$$\Pi_i^{\Lambda}(\tilde{p}^{\Lambda}, \tilde{T}^{\Lambda}) \leq \Pi_i^{\Lambda}(p^*, T^{\Lambda}) + \epsilon^{\Lambda}, \tag{19}$$

for any sequence of prices and service levels $\{(\tilde{p}^{\Lambda}, \tilde{T}^{\Lambda}),\ \Lambda \geq 0\}$ as long as $T^{\Lambda} \to 0$. Here ϵ^{Λ} is a sequence such that $\epsilon^{\Lambda}/\Lambda \to 0$. In particular, let $(\tilde{p}^{*,\Lambda}, \tilde{T}^{*,\Lambda})$ be the true optimal decision for this monopolist when the market scale is Λ. Assuming such a solution exists, equation (19) implies that

$$\frac{\Pi_i^{\Lambda}(p^{*,\Lambda}, T^{*,\Lambda}) - \Pi_i^{\Lambda}(p^*, T^{\Lambda})}{\Lambda} \to 0 \text{ as } \Lambda \to \infty.$$

This is an instance of the standard notion of fluid-scale asymptotic optimality. Hence, our results with respect to the fluid game are the game theoretic version of the fluid-scale asymptotic optimality for monopolists. In the same spirit, our equilibrium results for the diffusion game in §6 are a generalization of the diffusion-level asymptotic optimality results for monopolists. ■

Remark 5.5 (**service-level differentiation**) A fundamental implication of Theorem 5.2 is that the market is in an approximate equilibria if all firms set their prices according to p^* and choose a common (but very good) service level. While there might be many plausible explanations for the use of *industry standards*, Theorem 5.2 provides one such explanation in that it shows that following industry standards is not an irrational choice for firms competing on service levels and prices. In particular, firms need not significantly differentiate themselves in terms of service-level. The result can be interpreted as a one-sided decoupling result between prices and service levels (at least at the first order). The companies may set their prices according to p^*. Once the prices are fixed, a firm can exploit its large-scale efficiency and, in particular, the relative low cost of the service-based capacity, to match the service level of the competitor without moving significantly away from the equilibrium. ■

Note that Theorems 5.2 and 5.3 require only Assumptions 3.1 and 5.1. Most standard demand models, however, satisfy also Assumption 4.1 which, when imposed, allows us to improve on the convergence results in these theorems. To this end, recall that

$$r_i^{\Lambda} := \max\left\{\frac{1}{\Lambda^{\frac{1}{1+\alpha_i}}}, \frac{1}{\sqrt{\Lambda}}\right\},$$

where $\alpha_i,\ i \in \mathcal{I}$ are as in Assumption 4.1.

Theorem 5.6 (**distance from the fluid game**) *Suppose that Assumptions 3.1, 4.1 and 5.1 hold. Then, there exists a sequence* $\epsilon^{\Lambda} = O(1/r_1^{\Lambda}, \ldots, 1/r_I^{\Lambda})$ *such that, for each* $\Lambda \geq 0$, $(p^*, 0)$ *is an* ϵ^{Λ}*-Nash equilibrium for the* Λ^{th} *market game. Moreover,*

$$T_i^{*,\Lambda}(p^*, 0) \sim r_i^{\Lambda},\ i \in \mathcal{I} : \alpha_i > 1, \tag{20}$$

$$T_i^{*,\Lambda}(p^*, 0) = O(r_i^{\Lambda}),\ i \in \mathcal{I} : \alpha_i = 1, \tag{21}$$

and

$$T_i^{*,\Lambda}(p^*, 0) = o(r_i^{\Lambda}),\ i \in \mathcal{I} : \alpha_i < 1. \tag{22}$$

Theorem 5.6 characterizes the best service-level responses to the fluid-game equilibrium $(p^*, 0)$ The first part of the theorem strengthens our results in Theorems 5.2 and 5.3 by providing a better estimate of the rate of convergence. Instead of allowing any ϵ^{Λ} that satisfies $\epsilon^{\Lambda}/\Lambda \to 0$, Theorem 5.6 characterizes the growth rate of ϵ^{Λ} as a function of r^{Λ}.

Unfortunately, Theorem 5.6 is restricted to bounds on the service-level deviations from the fluid equilibrium. To obtain bounds on the prices we will need to impose additional restrictions on the demand models in consideration. We now gradually introduce the concepts and conditions that are required for that purpose. To this end, given $p \in \mathcal{P}$, let $\psi_i(p_{-i})$ be a best response of player i (in the fluid game) to prices p_{-i} of the competitors. If the best response is not unique we arbitrarily (but consistently) choose one. Set

$$\psi(p) := (\psi_1(p_{-1}), \ldots, \psi_I(p_{-I})).$$

By Assumption 5.1, the vector p^* is the unique solution to $p^* - \psi(p^*) = 0$. One expects that if p is a point in which no firm i can significantly improve its profits by deviating from p_i, then p should be close to the unique equilibrium p^*. Combined, Lemmas 5.7 and 5.8 identify conditions under which this intuition is valid.

Lemma 5.7 *Suppose that the following two conditions hold:*

(C1) *for each* $i \in \mathcal{I}$ *and* $T \in \Theta$*, the demand function* $\lambda_i(p, T)$ *is twice continuously differentiable in* p_i*.*

(C2) *there exists* $\delta > 0$ *such that for all* $p \in \mathcal{P}$ *and* $i \in \mathcal{I}$,

$$\frac{\partial^2}{\partial^2 p_i}\bar{\Pi}_i^P(p) \leq -\delta.^6$$

Then, there exists $\bar{\epsilon} > 0$ *such that, if* $p \in \mathcal{P}$ *and* $\epsilon \in [0, \bar{\epsilon}]^I$ *satisfy*

$$\bar{\Pi}_i^P(\psi_i(p_{-i}), p_{-i}) - \bar{\Pi}_i^P(p) \leq \epsilon_i, \; i \in \mathcal{I}, \tag{23}$$

they also satisfy

$$|p - \psi(p)| \leq M\sqrt{\epsilon}, \tag{24}$$

for some constant $M > 0$ *that is independent of* p*.*

Lemma 5.7 implies that, under the proper conditions on the demand model, if p is a price vector such that no firm can increase its profit significantly (in the fluid game) by unilaterally deviating from p, then p should be close to a corresponding best response vector $\psi(p)$. Under some conditions one can take one step further and show that if a vector p is close to its best response vector $\psi(p)$ then it must be close to the unique equilibrium of the fluid game, p^* [7]. One such condition is the well known "Diagonal Dominance Condition" which requires that

(C3) there exists $C < 1$ such that

$$\sum_{k \in \mathcal{I}} \left| \frac{\partial}{\partial p_k}\psi_i(p_{-i}) \right| \leq C, \; p \in \mathcal{P}, \; i \in \mathcal{I}.$$

Condition (C3) is also a sufficient condition for uniqueness of equilibrium p^* for the fluid game (see e.g. Theorem 5 in Cachon and Netessine (2004)). Example 5.1 at the end of this section shows how this condition is verified by means of the implicit function Theorem for the case of Multinomial Logit demand model. Whenever condition (C3) holds we say that the fluid game is *linearly continuous*. The motivation for this name comes from the following lemma.

[6]This condition can be imposed directly on the demand function by requiring that $\frac{\partial}{\partial^2 p_i}\lambda_i(p, 0)(p_i - c_i - \gamma_i/\mu_i) + 2\frac{\partial}{\partial p_i}\lambda_i(p, 0) < -\delta$.

[7]Note that the question whether p that is close to $\psi(p)$ must be close to p^* is essentially a question about the solution to the set of (possibly non-linear) equations $p = \psi(p)$. Indeed, putting $F_i(p) := p_i - \psi_i(p)$, what we want is that, if the set of equations $F(p) = 0$ has a unique solution p^*, then any p that satisfies $\|F(p)\| \leq \epsilon$ will be close to p^* in a way that is, to some extent, proportional to ϵ. Conditions on the function $F(\cdot)$ that guarantee the validity of such statements appear, for example, in the literature on convergence of algorithms for the solution of non-linear equations; see e.g. Gould et al. (2002) and the references therein.

Lemma 5.8 (**linear-continuity of fluid game**) *Suppose that* (C1)-(C3) *hold. Then, for all* $\epsilon \in \mathbb{R}_+^I$ *small enough*

$$|p - \psi(p)| \leq \epsilon,$$

implies that

$$|p - p^*| \leq \frac{1}{1 - C}|B^{-1}\epsilon|,$$

for the invertible matrix B *with* $B_{ii} = 0$ *and* $B_{ij} = 1$ *for all* $i \neq j$*. Consequently, by Lemma 5.7, there exist constants* $M, \bar{\epsilon} > 0$ *such that,*

$$\bar{\Pi}_i^P(\psi_i(p_{-i}), p_{-i}) - \bar{\Pi}_i^P(p) \leq \epsilon_i,\ i \in \mathcal{I} \tag{25}$$

for $\epsilon \in [0, \bar{\epsilon}]^I$*, implies that*

$$|p - p^*| \leq C_1\sqrt{\epsilon}. \tag{26}$$

Lemma 5.8 complements Lemma 5.7 to show that, under the "Diagonal Dominance Condition", if $p \in \mathcal{P}$ is such that no firm (in the fluid game) can increase its profits significantly by unilaterally deviating from it, then p must be close to p^*. In other words, Lemma 5.8 shows that, in the fluid game and under conditions (C1)-(C3), bounds in payoff space–as in equation (25)–imply bounds in action space–as in equation (26).

In the appendix we provide a framework for the continuity of the fluid game which replaces condition (C3) with a more general condition. To keep the presentation of our results as clear as possible, we first restrict ourselves to cases in which (C3) holds. Having introduced conditions (C1)-(C3) we can now extend the bounds in Theorem 5.6 to include bounds on deviations in both service-level and price. The matrix B that is used in the statement of the theorem is as in Lemma 5.8.

Theorem 5.9 (**distance from the fluid game**) *Suppose that Assumptions 3.1, 4.1, and 5.1 hold. Then, there exists a sequence* $\epsilon^\Lambda = O(1/r_1^\Lambda, \ldots, 1/r_I^\Lambda)$ *such that, for each* $\Lambda \geq 0$*,* $(p^*, 0)$ *is an* ϵ^Λ*-Nash equilibrium for the* Λ^{th} *market game. Moreover,*

$$T_i^{*,\Lambda}(p^*, 0) \sim r_i^\Lambda,\ i \in \mathcal{I} : \alpha_i > 1, \tag{27}$$

$$T_i^{*,\Lambda}(p^*, 0) = O(r_i^\Lambda),\ i \in \mathcal{I} : \alpha_i = 1, \tag{28}$$

and

$$T_i^{*,\Lambda}(p^*, 0) = o(r_i^\Lambda),\ i \in \mathcal{I} : \alpha_i < 1. \tag{29}$$

If, in addition, (C1)-(C3) *hold then*

$$|p^{*,\Lambda}(p^*,0)-p^*| = O(B^{-1}\sqrt{\delta^\Lambda}),\ i\in\mathcal{I}, \tag{30}$$

with $\delta_i^\Lambda = \frac{1}{\Lambda r_i^\Lambda} + (r_i^\Lambda)^{\alpha_i}$.

Theorem 5.9 uses the linear continuity of the fluid model to relate the market game to the fluid game. In particular, the theorem adds to the previous results by establishing the price bounds in (30). Conditions (C1)-(C3) are crucial in the proof of this theorem. Specifically, using the bound on the service-levels in (27)-(29) we show that the best response prices $p^{*,\Lambda}(p^*,0)$ constitute a r^Λ-Nash equilibrium for the *fluid game*. Namely, that $|\bar{\Pi}_i(p^{*,\Lambda}(p^*,0)) - \bar{\Pi}_i(\psi(p^{*,\Lambda}(p^*,0)))| \leq r_i^\Lambda$ for all $i\in\mathcal{I}$. Conditions (C1)-(C3) together with Lemma 5.8 are then used to obtain the price bounds.

Evidently then, the second part of Theorem 5.9 builds heavily on condition (C3) in obtaining the price bounds from the service-level bounds. Fortunately, various demand models satisfy (C3). We end this section with one such example.

Example 5.1 (**the ML model**) Fix $i\in\mathcal{I}$ with the demand model given in (13). The fluid game is, then, a game with demand functions

$$\lambda_i^P(p) := \lambda_i(p,0) = \frac{e^{a_i(0)-b_ip_i}}{v_0+\sum_{j\in\mathcal{I}} e^{a_j(0)-b_jp_j}},\ i\in\mathcal{I},$$

and payoff functions

$$\bar{\Pi}_i^P(p) := \frac{e^{a_i(0)-b_ip_i}}{v_0+\sum_{j\in\mathcal{I}} e^{a_j(0)-b_jp_j}}\left(p_i - c_i - \frac{\gamma_i}{\mu_i}\right).$$

Given p_{-i} the best response for firm i satisfies the equation

$$\left(1-\lambda_i^P(p_{-i},\psi_i(p_{-i}))\right)\left(\psi_i(p_{-i}) - c_i - \frac{\gamma_i}{\mu_i}\right) = \frac{1}{b_i}. \tag{31}$$

In particular, as there exists $\epsilon>0$ such that $\lambda_i^P(p) < 1-\epsilon$ for all $p\in\mathcal{P}$, we have that there exists δ such that $\psi_i(p_{-i}) > c_i + \frac{\gamma_i}{\mu_i} + \delta$ for all p_{-i}. Using the implicit function theorem and differentiating (31) with respect to p_j for $j\neq i$, we get

$$\begin{aligned} &-\frac{\partial}{\partial p_j}\lambda_i^P(p_{-i},\psi_i(p_{-i}))\left(\psi_i(p_{-i}) - c_i - \frac{\gamma_i}{\mu_i}\right) \\ &\quad = \frac{\partial}{\partial p_j}\psi_i(p_{-i})\left(\frac{\partial}{\partial p_i}\lambda_i^P(p_{-i},\psi_i(p_{-i}))\left(\psi_i(p_{-i}) - c_i - \frac{\gamma_i}{\mu_i}\right) - (1-\lambda_i^P(p_{-i},\psi_i(p_{-i})))\right), \end{aligned} \tag{32}$$

where $\frac{\partial}{\partial p_j}\lambda_i^P(p_{-i}, \psi_i(p_{-i}))$ and $\frac{\partial}{\partial p_i}\lambda_i^P(p_{-i}, \psi_i(p_{-i}))$ are the partial derivatives with respect to p_j and p_i respectively at the point $p = (p_{-i}, \psi(p_{-i}))$. Plugging (31) into (32) as well as the derivatives of $\lambda_i^P(p)$ with respect to p_i and p_j we have that

$$\sum_{j\in\mathcal{I}} \left|\frac{\partial}{\partial p_j}\psi_i(p_{-i})\right| = \sum_{j\in\mathcal{I}} \left|\frac{b_j\lambda_i^P(p_{-i}, \psi_i(p_{-i}))\lambda_j^P(p_{-i}, \psi_i(p_{-i}))}{b_i(1-\lambda_i^P(p_{-i}, \psi_i(p_{-i})))}\right|.$$

The right hand side is strictly less than 1 provided that

$$\frac{\sum_{j\in\mathcal{I}} b_j\lambda_j^P(p)}{b_i} < 1,\ i \in \mathcal{I},\ p \in \mathcal{P}. \tag{33}$$

This condition is the "Dominant Diagonal Condition" for the ML model. Hence, condition (C3) holds for the ML model provided that (33) holds. By differentiating $\bar{\Pi}_i^P(p)$ twice one can also verify that (C1) and (C2) hold for the ML model. ■

This concludes the analysis of the fluid game and we turn to the introduction and analysis of the diffusion game.

6. A diffusion game

In this section we improve on our understanding from the fluid game by introducing and studying a diffusion game. This game is more refined than the fluid game and can be interpreted as a second-order approximation for the market game. The sequence of diffusion games is constructed by replacing the service facilities (the $M/M/N$ queues) by their diffusion approximation. The challenge is then to relate the equilibria of the diffusion game to those of the original market game. We will show that the diffusion game provides an equilibrium characterization that is asymptotically correct in diffusion scale–thus paralleling the diffusion-scale asymptotic optimality results for monopolists; see e.g. Borst et al. (2004). The bounds in payoff space and action space that come out of this analysis will allow us to characterize some aspects of the market behavior.

6.1 Definition and characterization

In constructing the diffusion game, we use Lemma 6.1 below to replace the service-based capacity with a simple expression. The lemma relies on Borst et al. (2004) in approximating the delay distribution by an expression that uses the asymptotic version, $\mathbf{P}(\cdot)$, for the probability of delay as identified in Halfin and Whitt (1981). In the following and hereafter, we put $R_i(p,T) := \Lambda_i(p,T)/\mu_i$ for $(p,T) \in \mathcal{P} \times \Theta$.

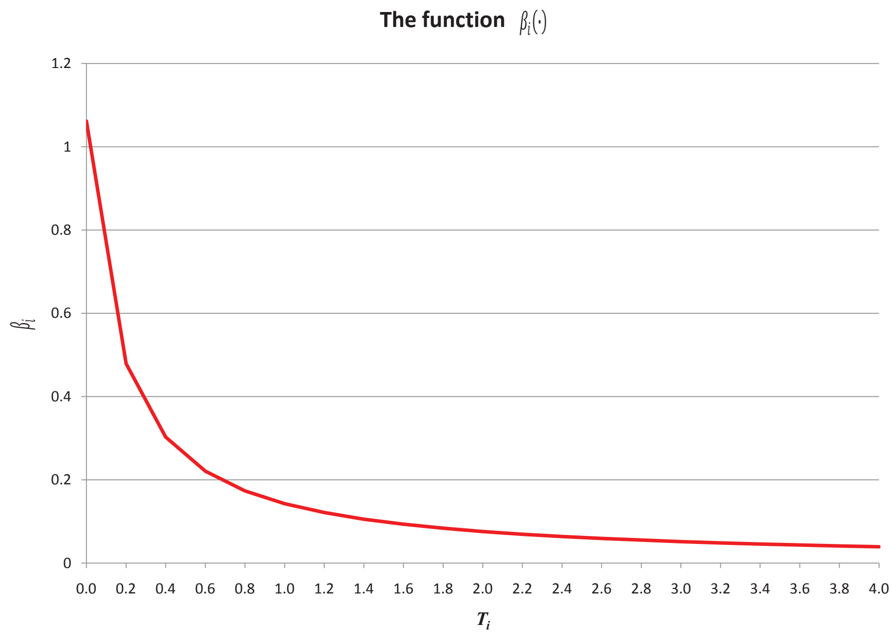

Figure 1: The function $\beta_i(\cdot)$ for (p, T) such that $R_i(p, T) = 100$ and $\mu_i = 1$

Lemma 6.1 (**M/M/N Lemma**) *Fix a sequence* $\{(p^\Lambda, T^\Lambda),\ \Lambda \geq 0\}$ *such that, for each* Λ, $(p^\Lambda, T^\Lambda) \in \mathcal{P} \times \Theta$. *Then, for all* $i \in \mathcal{I}$,

$$\begin{aligned}\hat{e}_i(\Lambda_i, T_i^\Lambda) &= \beta_i(\sqrt{R_i(p^\Lambda, T^\Lambda)}T_i^\Lambda)\sqrt{R_i(p^\Lambda, T^\Lambda)} \\ &+ o\left(\beta_i(\sqrt{R_i(p^\Lambda, T^\Lambda)}T_i^\Lambda)\sqrt{R_i(p^\Lambda, T^\Lambda)}\right).\end{aligned}$$

where, given (p, T), β_i *is the unique solution to*

$$\mathbf{P}(x)e^{-\mu_i x\sqrt{R_i(p,T)}T_i} = \phi.$$

Here

$$\mathbf{P}(x) = \left[1 + \frac{xZ(x)}{z(x)}\right]^{-1},$$

where $z(\cdot)$ *and* $Z(\cdot)$ *are, respectively, the standard normal density function and its cumulative distribution function. Furthermore, the function* $\beta_i(\cdot)$ *is a continuously differentiable and convex decreasing function on* $[0, \bar{T}]$.

Lemma 6.1 states that, given $(p, T) \in \mathcal{P} \times \Theta$, the service-based capacity can be written as the sum of $\beta_i(\sqrt{R_i(p, T)}T_i)\sqrt{R_i(p, T)}$ and a smaller order term so that the profit functions can be written as follows:

$$\begin{aligned}\Pi_i^\Lambda(p,T) \;&:=\; \Lambda_i(p,T)\left(p_i - c_i - \frac{\gamma_i}{\mu_i}\right) - \gamma_i\beta_i(\sqrt{R_i(p,T)}T_i)\sqrt{R_i(p,T)} \\ &+\; o\left(\beta(\sqrt{R_i(p,T)T_i})\sqrt{R_i(p,T)}\right). \end{aligned} \tag{34}$$

Even without the smaller order term, the expressions in (34) are complex functions due to the dependence of $\sqrt{R_i(\cdot,\cdot)}$ term on the entire vector (p,T). By Theorem 5.3 we know, however, that a sequence $\{(p^\Lambda,T^\Lambda),\Lambda \geq 0\}$ of approximate Nash equilibria must satisfy that $(p^\Lambda,T^\Lambda) \to (p^*,0)$ and, by the assumed continuity of the demand functions in Assumption 3.1, that

$$\frac{\sqrt{R_i(p^\Lambda,T^\Lambda)} - \sqrt{R_i(p^*,0)}}{\sqrt{R_i(p^\Lambda,T^\Lambda)}} \to 0 \text{ as } \Lambda \to \infty. \tag{35}$$

These observations motivate the introduction of the *diffusion game* as an approximation for Λ^{th} market game for all Λ large enough.

Definition 6.2 (**the diffusion game**) *Fix $\Lambda \geq 0$. The Λ^{th} diffusion game has I players, profit functions*

$$\hat{\Pi}_i^\Lambda(p,T) := \Lambda_i(p,T)\left(p_i - c_i - \frac{\gamma_i}{\mu_i}\right) - \gamma_i\beta_i(\sqrt{R_i(p^*,0)}T_i)\sqrt{R_i(p^*,0)},\; i \in \mathcal{I},$$

and strategy space $\mathcal{P}\times\Theta$.

Note that, in defining the new profit function $\hat{\Pi}_i^\Lambda(\cdot,\cdot)$, we have replaced the service-based capacity, $\hat{e}_i(\cdot,\cdot)$, by a simpler term that depends on the price equilibrium of the fluid game, p^*, but is otherwise independent of the actual price vector p and of the service-levels of the competitors as in T_{-i}. Moreover, this term is convex and continuous in T_i. This relative simplicity renders the diffusion game tractable for Nash equilibrium analysis in some cases. For example, it suffices to require that, for each $i \in \mathcal{I}$, the demand function $\lambda_i(p,T)$ is jointly concave in the decision (p_i,T_i) of firm i. For future reference we assign this condition a number.

(C4) for each $i \in \mathcal{I}$, and each (p_{-i},T_{-i}), the demand function $\Lambda_i(p,T)$ is jointly concave in (p_i,T_i).[8]

[8] When the demand functions are twice continuously differentiable, it can be easily verified that the concavity of $\lambda_i(p,T)$ in (p_i,T_i) and the monotonicity assumed in Assumption 3.1 imply the concavity of the function $\lambda_i(p,T)(p_i - c_i - \gamma_i/\mu_i)$.

6.2 The quality of the approximation

We now analyze the quality of the diffusion game as an approximation for the market game. It turns out that, in order to improve on the quality of the approximation of the fluid game (as given in Theorem 5.9) we need to strengthen Assumption 4.1 as follows:

(C5) there exists $0 < \delta \leq \bar{T}$ and a continuous function $f_i(\cdot,\cdot) : \mathcal{P} \times [0,\bar{T}]^{I-1} \to \mathbb{R}$ such that

$$\lim_{x \to 0} \frac{\lambda_i(p, T_{-i}, x) - \lambda_i(p, T_{-i}, 0)}{x^{\alpha_i}} \to f_i(p, T_{-i})$$

for every $(p, T_{-i}) \in \mathcal{P} \times [0,\delta]^{I-1}$.

Note that condition (C5) implies in particular Assumption 4.1. This condition holds for various demand models; see Example 6.1 below and the linear demand model in §7. We then have the following result, which is followed by an intuitive explanation on the role of (C5). The matrix B that is used in the statement of the theorem is as in Lemma 5.8–it is the $I \times I$ matrix given by $B_{ii} = 0,\ i = 1, \ldots, I$ and $B_{ij} = 0$ for all $i \neq j$. Also, we recall that

$$r_i^\Lambda := \max \left\{ \frac{1}{\Lambda^{\frac{1}{1+\alpha_i}}}, \frac{1}{\sqrt{\Lambda}} \right\},$$

as constructed in §4.

Theorem 6.3 (**distance from the diffusion game**) *Suppose that Assumptions 3.1 and 5.1 hold in addition to conditions* (C1)-(C5)*, and let* $\{(p^\Lambda, T^\Lambda),\ \Lambda \geq 0\}$ *be a sequence such that* (p^Λ, T^Λ) *is a Nash equilibrium for the* Λ^{th} *diffusion game. Then, there exists a sequence* $\epsilon^\Lambda = o\left(1/r_1^\Lambda, \ldots, 1/r_I^\Lambda\right)$, *such that* (p^Λ, T^Λ) *is an* ϵ^Λ*-Nash equilibrium for the* Λ^{th} *market game. Moreover,*

$$T_i^{*,\Lambda}(p^\Lambda, T^\Lambda) = T_i^\Lambda + o(r_i^\Lambda), \quad \textit{and} \quad p_i^{*,\Lambda}(p^\Lambda, T^\Lambda) = p_i^\Lambda + o\left(B^{-1}\sqrt{\zeta^\Lambda}\right),$$

where $\zeta_i^\Lambda = (r_i^\Lambda)^{\alpha_i}$.

To clarify the role of condition (C5) in Theorem 6.3, assume that a Nash equilibrium (p^Λ, T^Λ) does exist for the Λ^{th} market game and recall that, in that case,

$$T_i^\Lambda := \operatorname*{argmax}_{x \in [0,\bar{T}]} \left[\Lambda_i(p^\Lambda, T_{-i}^\Lambda, x) - \Lambda_i(p^\Lambda, T_{-i}^\Lambda, 0)\right] \left(p_i^\Lambda - c_i - \frac{\gamma_i}{\mu_i}\right) - \gamma_i \hat{e}_i(\Lambda_i, x).$$

From §4 we know that T_i^Λ will be close to 0, hence we may heuristically replace the "waiting cost" $\Lambda_i(p^\Lambda, T_{-i}^\Lambda, x) - \Lambda_i(p^\Lambda, T_{-i}^\Lambda, 0)$ by an approximation of its behavior around 0. Assumption

4.1 only guarantees that the behavior will be proportional to x^{α_i}. While this is sufficient to obtain the relatively crude results of Theorem 5.6, it is not sufficient for the finer characterization in Theorem 6.3 above. For this result, we need to identify the functional form of the behavior around $T = 0$. This is guaranteed by means of condition (C5). Theorem 6.3 implies that the service-level and price equilibrium, (p^Λ, T^Λ), of the diffusion game provide precise approximations for the real market outcomes, thus mimicking the role of the diffusion approximation in the context of monopolists. More precisely, by using the diffusion game to determine the firms' decisions, the compromise in profits is negligible with respect to the cost of the service-based capacity which is, in turn, proportional to $1/r_i^\Lambda$. This is indeed reminiscent of the notion of asymptotic optimality used in the context of monopolists.

Remark 6.4 (**the size of ϵ^Λ and diffusion-level asymptotic optimality**) Consider the case in which the set of firms $\mathcal{I}$ consists of a single firm, firm 1–a monopolist. An equilibrium of the diffusion game is then a maximizer of $\hat{\Pi}_1^\Lambda(p_1, T_1)$ where, $\hat{\Pi}_1^\Lambda(\cdot,\cdot)$ is the profit function in Definition 6.2. Pick

$$(p_1^\Lambda, T_1^\Lambda) \in \operatorname*{argmax}_{p,T} \hat{\Pi}_1^\Lambda(p, T),$$

i.e $(p_1^\Lambda, T_1^\Lambda)$ is a maximizer of the diffusion-game profit when the market scale is Λ. Then, the ϵ^Λ-Nash equilibria result in Theorem 6.3 reduces, in the monopolist setting, to asymptotic optimality in the sense of Borst et al. (2004). Specifically, Theorem 6.3 implies for this setting that, for any sequence $\{(\tilde{p}_1^\Lambda, \tilde{T}_1^\Lambda),\ \Lambda \geq 0\}$ of prices and service levels

$$\liminf_{\Lambda\to\infty} \frac{\Pi_1^\Lambda(p_1^\Lambda, T_1^\Lambda) - \Pi_1^\Lambda(\tilde{p}_1^\Lambda, \tilde{T}_1^\Lambda)}{1/r_1^\Lambda} \geq 0,$$

where $\Pi_1^\Lambda(\cdot,\cdot)$ is the profit function in the Λ^{th} market game; see Definition 3.1. In other words the optimality gap for this monopolist, if it chooses to use the outcome of the diffusion game, is of the order of $o(1/r_1^\Lambda)$. If the monopolist has $r_1^\Lambda = 1/\sqrt{\Lambda}$, then the optimality gap is $o(\sqrt{\Lambda})$ which corresponds to the prevalent optimality gap in the literature that considers asymptotic optimality in the Halfin-Whitt regime.

We emphasize that our main result is stronger than asymptotic optimality. We not only provide bounds on the optimality gap with respect to profits but also with respect to the price and service-level decisions. Theorem 6.3 shows that, with $(\tilde{p}_1^\Lambda, \tilde{T}_1^\Lambda)$ being an *optimal* solution for the monopolist when the market scale is Λ, then the sequence $\{(\tilde{p}_1^\Lambda, \tilde{T}_1^\Lambda), \Lambda \geq 0\}$ must satisfy

$$\tilde{T}_1^\Lambda = T_1^\Lambda + o(r_1^\Lambda) \text{ and } \tilde{p}_1^\Lambda = p_1^\Lambda + o(B^{-1}\sqrt{\zeta_1^\Lambda}),$$

where the vector $(p_1^\Lambda, T_1^\Lambda)$ is the optimal solution to the diffusion game. ■

It turns out that, under the conditions of Theorem 6.3 we can be more precise about the service-level characterization. The following lemma shows that the service-level choice under the diffusion-game equilibrium can be characterized in close form up to an error of size $o(r_i^\Lambda)$.

Lemma 6.5 *Suppose that Assumptions 3.1 and 5.1 hold in addition to condition* (C5)*, and let* $\{(p^\Lambda, T^\Lambda), \Lambda \geq 0\}$ *be any sequence such that* $(p^\Lambda, T^\Lambda) \to (p^*, 0)$ *as* $\Lambda \to \infty$*. Then, for each* $i \in \mathcal{I}$*,*

$$\frac{T_i^{*,\Lambda}(p^\Lambda, T^\Lambda)}{r_i^\Lambda} \to \eta_i^* \text{ as } \Lambda \to \infty,$$

where $\eta_i^* = 0$ *if* $\alpha_i < 1$ *and it equals*

$$\eta_i^* = \underset{\eta \geq 0}{\operatorname{argmax}}\, \eta^{\alpha_i} f_i(p^*, 0) \left(p_i^* - c_i - \frac{\gamma_i}{\mu_i} \right) - \gamma_i \beta_i(\eta_i) \left(\frac{\lambda_i(p^*, 0)}{\mu_i} \right)^{\frac{1}{\alpha_i + 1}},$$

if $\alpha_i \geq 1$*. Here* $\beta_i(\eta_i)$ *is the solution* x *to* $\mathbf{P}(x) e^{-\mu_i x \sqrt{\frac{\lambda_i(p^*,0)}{\mu_i}}} = \phi$ *whenever* $\alpha_i = 1$ *and is the solution to* $e^{-\mu_i x \left(\frac{\lambda_i(p^*,0)}{\mu_i}\right)^{\frac{1}{\alpha_i+1}}} = \phi$ *when* $\alpha_i > 1$.

Remark 6.6 (**hierarchical decoupling**) Combined, Lemma 6.5 and Theorem 6.7 justify referring to demand models that satisfy (C5) as demand models that admit a *hierarchical decoupling*. Indeed, Lemma 6.5 shows that service-level choices depend on the actions of its competitors mostly through their prices (and not their service levels). Moreover, they depend on these prices only through their fluid game equilibrium p^*. Practically, this suggests that service level and price choices can be made in a sequential manner rather than jointly. The firms will first choose their price based on the fluid game, i.e. disregarding service level considerations. Based on the these prices the firms will make their service-level choices. While the firms might choose to adjust their prices at a later stage in response to the actions of the competition, they will not need to revisit their service-level choices. These can remain fixed without any significant compromise to the firm's profits. ■

Lemma 6.5 allows us to go one step further. It allows us to replace condition (C4)–that guarantees the existence of a Nash equilibrium for the diffusion game, with a condition that is imposed on a much simpler "perturbed" fluid game. To this end, let *the fluid game on T* be the I player game with profit functions

$$\bar{\Pi}_i^{T,P}(p) := \lambda_i(p, T) \left(p_i - c_i - \frac{\gamma_i}{\mu_i} \right),\ i \in \mathcal{I},$$

and strategy space $\mathcal{P}$. The fluid game on $T = 0$ is the fluid game from Definition 5.1. Theorem 6.7 below provides a characterization of the ϵ^Λ Nash equilibrium in terms of the Nash equilibrium of the fluid game on $\eta^\Lambda := (\eta_1^* r_1^\Lambda, \ldots, \eta_I^* r_I^\Lambda)$ with $(\eta_1^*, \ldots, \eta_I^*)$ as in Lemma 6.5. Note that, in contrast to Theorem 6.3, here we do not impose condition (C4). Instead, we assume uniqueness of equilibrium for the fluid game on η^Λ.

Theorem 6.7 *Suppose that Assumptions 3.1 and 5.1 hold in addition to conditions* (C1)-(C3) *and* (C5). *Assume that, for all* Λ *large enough, the fluid game on* $\eta^\Lambda := (\eta_1^* r_1^\Lambda, \ldots, \eta_I^* r_I^\Lambda)$ *has a unique Nash equilibrium* $p^\Lambda(\eta^\Lambda)$. *Then, there exists a sequence* $\epsilon^\Lambda = o\left(1/r_1^\Lambda, \ldots, 1/r_I^\Lambda\right)$, *such that* $(p^\Lambda(\eta^\Lambda), \eta^\Lambda)$ *is an* ϵ^Λ*-Nash equilibrium for the* Λ^{th} *market game. Moreover,*

$$T_i^{*,\Lambda}(p^\Lambda(\eta^\Lambda), \eta^\Lambda) = \eta_i^\Lambda + o(r_i^\Lambda), \quad \text{and} \quad p_i^{*,\Lambda}(p^\Lambda(\eta^\Lambda), \eta^\Lambda) = p_i^\Lambda(\eta^\Lambda) + o\left(B^{-1}\sqrt{\zeta^\Lambda}\right),$$

where $\zeta_i^\Lambda = (r_i^\Lambda)^{\alpha_i}$.

Example 6.1 (**back to the Multinomial-Logit demand**) By Example 5.1, (C1)-(C3) all hold provided that

$$\frac{\sum_{j \in \mathcal{I}} b_j \lambda_j^P(p)}{b_i} < 1, \; i \in \mathcal{I}, \; p \in \mathcal{P}. \tag{36}$$

It remains to show that (C5) holds and that a unique equilibrium exists for the fluid game on η^Λ. First, we claim that (C5) holds with

$$f_i(p, T_{-i}) := \frac{k_i v_i(p_i, 0)(1 - \lambda_i(p, T_{-i}, 0))}{1 + \sum_{j \neq i} v_j(p_j, T_j) + v_i(p_i, 0)}.$$

The simple, but detailed, argument is given in the e-companion. The proof can be useful as a guideline towards the verification of (C5) for other demand models. It can be verified that for the ML demand model, and for all T small enough, the corresponding fluid game on T has a unique equilibrium. Indeed, provided the fluid game on $T = 0$ satisfies condition (C3), the fluid game on T (for T in a sufficiently small neighborhood of 0) will satisfy condition (C3) by virtue of the continuity of the best response functions and their derivatives. Condition (C3), in turn, guarantees the uniqueness of equilibria for that game. Since $\eta^\Lambda \to 0$ as $\Lambda \to \infty$, we will have that the ML demand model satisfies the conditions of Theorem 6.7. ■

Remark 6.8 (**level of sub-optimality and freedom in pricing**) Interpreting ϵ_i^Λ as the level of sub-optimality for firm i (see Remarks 5.4 and 6.4), Theorem 6.3 provides an insight into the relations between the operational regime of a firm and its pricing decision under a given sub-optimality level. The theorem states that the sub-optimality level of $o(1/r_i^\Lambda)$ is preserved as long as the price distance from the diffusion-game equilibrium price p_1^Λ is of order $o((r_i^\Lambda)^{\alpha_i})$ where α_i is the exponent from Assumption 4.1; see §4.

As an example, assume that (C4) holds and let $\{(p^\Lambda, T^\Lambda),\ \Lambda \geq 0\}$ be a sequence of Nash equilibria for the diffusion game. Assume that firm i has $r_i^\Lambda = 1/\sqrt{\Lambda}$ and, consequently, operates optimally in the QED regime. By theorem 6.3, this firm can choose to charge the price $\tilde{p_i}^\Lambda = p_i^\Lambda + 1/\Lambda^{1/5}$ without compromising its level of optimality–which will remain $o(\sqrt{\Lambda})$. In contrast, assume that firm i has $r_i^\Lambda = 1/\Lambda^{1/3} > 1/\sqrt{\Lambda}$ so that it operates optimally in the ED regime. Then, if it deviates from the diffusion-game price equilibrium by the same $1/\Lambda^{1/5}$, its compromise in profits will now be order-of-magnitude greater than $o(1/r_i^\Lambda)$. In other words, a firm operating in the QED regime has a larger interval from which it can choose its prices without compromising its level of optimality. ■

Remark 6.9 (**global stability**) The results stated in this section focus on the existence and, to some extent, uniqueness of approximate equilibria. Accordingly, in the spirit of equilibrium analysis, the focus is on unilateral deviations. In section C of the appendix we strengthen these results by proving a global stability result. We show that, for any starting point $(p, T) \in \mathcal{P} \times \Theta$, the market converges to a neighborhood of the diffusion game equilibrium and this neighborhood is exactly the one characterized in Theorem 6.7. ■

7. Example: A linear demand model

In this section we provide a numerical example to illustrate the approximations in §6. The example is based on the linear demand model in Allon and Federgruen (2007). Specifically, we use demand functions specified by:

$$\lambda_i(p,T) = \left[a_i(T_i) - b_i p_i + \sum_{j \neq i} a_{ij}(T_j) + \sum_{j \neq i} l_{ij} p_j\right]^+, \tag{37}$$

and we assume that

$$a_i(T_i) = a_i - k_i T_i \text{ and } a_{ij}(T_j) = k_{ij} T_j \tag{38}$$

for strictly positive constants $a_i, k_i, k_{ij},\ i, j \in \mathcal{I}$. We further assume that a *uniform* price increase by all I firms cannot result in an increase in any firm's demand volume and that a price increase by a given firm cannot result in an increase of the industry's aggregate demand volume, i.e.,

$$(D) \quad b_i > \sum_{j \neq i} l_{ij}, i = 1, \ldots, I; \qquad (D') \quad b_i > \sum_{j \neq i} l_{ji}, i = 1, \ldots, I. \tag{39}$$

The requirements in (39) guarantee that the ”Dominant Diagonal” condition–condition (C3)–holds for the fluid game of the linear demand model. Equation (39) guarantees that the $I \times I$ matrix A, defined by $A_{ii} = 2b_i$ and $A_{ij} = -l_{ij}$ for $i \neq j$, is invertible. We let A^{-1} be its inverse. Finally, we make the assumption that $a_i(0) + \sum_{j \neq i} l_{ij} p_j^{min} > b_i p_i^{min}$ so that

$$\lambda_i(p, T) > 0, \forall (p, T) \in \mathcal{P} \times \Theta. \tag{40}$$

The Nash equilibria of the fluid game for the linear demand model can be characterized in closed form. The fluid-game's best response function by firm i is given by

$$\psi_i(p_{-i}) = \frac{a_i(0) - \sum_{j \neq i} a_{ij}(0) + \sum_{j \neq i} l_{ij} p_j + b_i \left(c_i + \frac{\gamma_i}{\mu_i} \right)}{2b_i},$$

and the invertibility of A then guarantees that the unique solution to the system $p = \psi(p)$ is the unique Nash equilibrium of the fluid game. We denote this equilibrium by p^*.

Using (38), it is easily verified that Assumption 4.1 holds for this linear demand model with the exponents, $\{\alpha_i, \ i \in \mathcal{I}\}$, in that assumption being all equal to one. In particular, we have that $r_i^\Lambda = 1/\sqrt{\Lambda}$ for all $i \in \mathcal{I}$ so that, by Theorem 5.6 and Theorem 4.2, all firms operate optimally in the QED regime. It can be also easily verified that conditions (C1)-(C5) hold for the linear demand model and, consequently, that it satisfies the conditions of Theorem 6.3. The Nash equilibria of the corresponding diffusion game have a simple, closed-form, characterization. Specifically, an equilibrium (p, T) of the corresponding diffusion game must satisfy

$$(Ap)_i = a_i - k_i T_i + \sum_{j \neq i} k_{ij} T_j + b_i \left(c_i + \frac{\gamma_i}{\mu_i} \right), \tag{41}$$

and

$$-\left(p_i - c_i - \frac{\gamma_i}{\mu_i} \right) k_i - \gamma_i \beta_i'(\sqrt{R_i(p^*, 0)} T_i) \sqrt{R_i(p^*, 0)} = 0. \tag{42}$$

Theorem 7.1 (**the diffusion game of the linear demand model**) *For each Λ, a Nash equilibrium exists for the Λ^{th} diffusion game. Let $\{(p^\Lambda, T^\Lambda), \ \Lambda \geq 0\}$ be a sequence of such equilibria. Then,*

$$p_i^\Lambda = p_i^* + \frac{\varrho_i}{\sqrt{R_i(p^*, 0)}}, \ i \in \mathcal{I}, \tag{43}$$

and

$$T_i^\Lambda = \frac{\eta_i}{\sqrt{R_i(p^*, 0)}}, \ i \in \mathcal{I}. \tag{44}$$

Here, $\varrho := (\varrho_1, \ldots, \varrho_I)$ *and* $\eta := (\eta_1, \ldots, \eta_I)$ *are the unique solution to the system of equations*

$$(A\varrho)_i = k_i\eta_i - \sum_{j\neq i} k_{i,j}\eta_j, \; i \in \mathcal{I}, \tag{45}$$

$$\gamma_i\beta_i'(\eta_i) = k_i\frac{\mu_i}{\lambda_i(p^*,0)}\left(p_i^* - c_i - \frac{\gamma_i}{\mu_i}\right), i \in \mathcal{I}, \tag{46}$$

where (given η_i*)* $\beta_i(\eta_i)$ *is the unique solution of*

$$\mathbf{P}(\beta_i(\eta_i))e^{-\mu_i\beta_i(\eta_i)\eta_i} = \phi.$$

The linear demand model hence provides a clean illustration of the results developed in sections 4-6 as it allows, for example, to characterize in explicit and simple terms the diffusion game equilibria. It serves also to illustrate the result in Lemma 6.5. Indeed, we see in Theorem 7.1 that the diffusion game service-levels and price are related only through the fluid game price equilibrium.

Numerical example: We consider an industry with $I = 3$ firms, $\bar{T} = 1$, and cost parameters $c_1 = c_2 = 20, c_3 = 5$, and $\gamma_1 = \gamma_2 = 35, \gamma_3 = 50$. This setting can be interpreted as having firm 3 to be 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 the capacity costs, γ, are lower but the per-customer access cost, c, is higher. We assume that all firms experience identical price sensitivities. Specifically, we assume that $a_1 = a_2 = 2.05$ and $a_3 = 2.95$. We also set $b_i = 1$ for all $i = 1, 2, 3$ and $l_{i,j} = 0.5$ for all $i \neq j$. Finally, we set $\kappa_i = 4/3$ for all $i = 1, 2, 3$ and $\kappa_{ij} = 2/300$. The results are depicted in Figure 2.

Figure 2 is constructed as follows: we first solve the first order conditions (45) and (46) to obtain the vectors η and ϱ. We then use (43) and (44) to construct the sequence of diffusion-game Nash equilibria $\{(p^\Lambda, T^\Lambda), \Lambda \geq 0\}$. Then for each fixed Λ, we initialize the Λ^{th} **market game** at the point (p^Λ, T^Λ) and find best responses $T_i^{*,\Lambda}(p^\Lambda, T^\Lambda)$ and $p_i^{*,\Lambda}(p^\Lambda, T^\Lambda)$ for each firm $i = 1, 2, 3$.

Figure 2 displays the maximal profitable deviations for firm 1 (the quality of the approximations is similar for the other two firms). The left-hand graph corresponds to deviations in the service-level dimension. Specifically, the solid series depicts the sequence of the service-level choice in the diffusion game equilibrium, T_1^Λ, as a function of Λ. For each value of Λ we calculate the best response in the Λ^{th} market game, $T_1^{*,\Lambda}(p^\Lambda, T^\Lambda)$. The pointed and dashed series are then, respectively, the upper and lower bounds that this best-responses induce, i.e, the Λ^{th} point in the

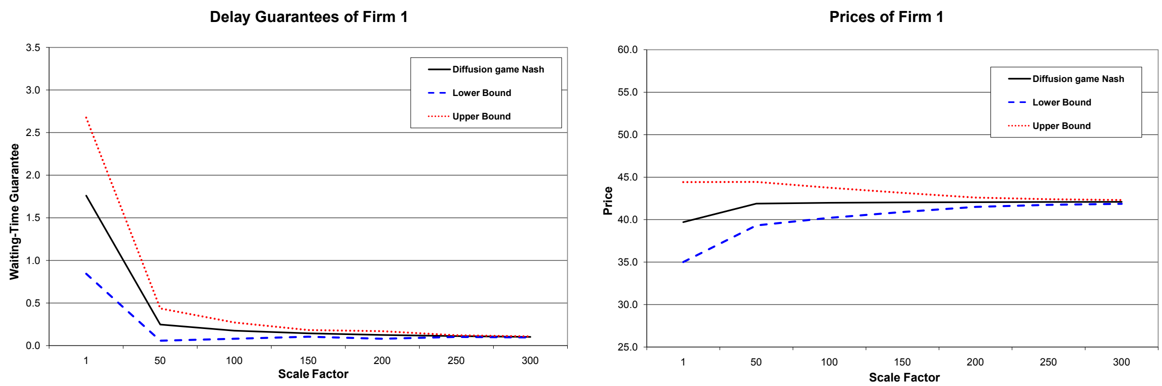

Figure 2: The quality of the approximation – another example of the linear model

pointed series corresponds to the value $T_i^\Lambda + |T_i^{*,\Lambda}(p^\Lambda, T^\Lambda) - T_i^\Lambda|$ and the Λ^{th} point in the dashed series to the value $T_i^\Lambda - |T_i^\Lambda - x^\Lambda|$. The fact that the pointed and dashed lines are very close to T_i^Λ is the numerical illustration of our result in Theorem 6.3 that $T_i^{*,\Lambda}(p^\Lambda, T^\Lambda) = T_i^\Lambda + o(T_i^\Lambda)$. The right hand graph then repeats the same steps for the sequence of prices p_i^Λ obtained from the diffusion game and the corresponding best response sequence $\{p_i^{*,\Lambda}(p^\Lambda, T^\Lambda),\ \Lambda \geq 0\}$.

We see, then, that profitable deviations from the Nash equilibrium of the diffusion game (p^Λ, T^Λ) are small for both the service-level and the price choices. Notably, in this example the prices of the diffusion game are getting closer to the diffusion game prices as Λ grows.

8. Discussion

In this paper we study markets with multiple large-scale service providers. To do so, we develop a novel framework that combines the notions of ϵ-Nash equilibrium, market replication and heavy-traffic to study market equilibria. The ϵ-Nash framework allows us to go beyond the scope of models of competition for which Nash equilibrium exists and use relatively general demand and capacity models. The notion of market replication allows us to discuss trends in terms of stability and market outcomes in sequences of markets such, as the impact of the market scale on the interdependence between the pricing and service-level decisions. Combined with the notion of heavy-traffic, which is well studied for monopolists, this framework allows us to characterize equilibrium behavior and obtain insights for markets in which Nash equilibrium need not necessarily exist.

The framework developed in this paper can be applied to other competitive settings in which congestion and queueing play important roles. The framework is especially relevant in settings that satisfy two conditions: (a) a Nash equilibrium does not exist or is intractable for characterization, and (b) there are available approximations for the underlying queueing systems that can be used to construct a tractable approximate game. In our setting, the diffusion game–which is based on many-server heavy-traffic approximations–plays the role of this approximate game, but this need not be the case.

Indeed, one can apply the same approach to markets with single-server suppliers in which the service rate, rather than the number of servers, is the capacity decision variable. In these cases, we expect that the so-called *conventional heavy-traffic* approximations – in which the number of servers is kept fixed and the load approaches one–would play a key role in supplying the approximations that would replace each of the suppliers in the construction of the diffusion game. In these single-server settings our approach can be used to study demand models in which the customers are sensitive to the whole sojourn time rather than solely to the waiting time in queue.

To illustrate this latter claim, we consider a market with two firms such that firm i faces the following logit demand

$$\lambda_i(f_1, f_2) = m\frac{a_i e^{bf_i}}{v_0 + a_1 e^{b_1 f_1} + a_2 e^{b_2 f_2}}.$$

Here m plays the role of the market size. Also, f_i is the *full price* “charged” by firm i. That is $f_i = p_i + s_i$ where s_i is the average sojourn time of customers served by firm i. Hence, rather than treating price and service level as independent attributes, this game will consider only their linear combination. Both firms operate through a single server facility so that firm i adjusts its service rate μ_i rather than the number of servers. The cost of capacity is then $c_i\mu_i$ for $c_i > 0$.

$$\Pi_i^m(f_1, f_2) = (f_i - c_i)\lambda_i - 2\sqrt{c_i\lambda_i}\ ,\ i = 1, 2,$$

where c_i is the cost of a unit of capacity. This model is very similar (apart from the existence of an outside option) to the one considered in Cachon and Harker (2002).

We use the parameters $c_1 = c_2 = 3.75$, $a_i = -b_i = 1$ for $i = 1, 2$, as in the example in Figure 3 of Cachon and Harker (2002). As in Cachon and Harker (2002), equilibrium does not exist when $m = 1$. For larger values of m, however, the market seems to have a Nash equilibrium. Still, this Nash equilibrium need not be unique. The fluid game allows us, however, to obtain a first-order approximation of the full-price equilibrium. Following our approach in this paper we first define a fluid game by removing the service-based capacity cost $2\sqrt{c_i\lambda_i}$. The m^{th} fluid game is the game with profit functions $\bar{\Pi}_i^m(f) = (f_i - c_i)\lambda_i^m(f_1, f_2)$. This fluid game does have the equilibrium

$f^* = (5.725, 5.725)$. Moreover, it can be easily verified the diagonal dominance condition in equation (33) holds for the above parameters, so that this equilibrium is the unique equilibrium of the fluid game.

We numerically compute equilibria for each value of m for the original game with profit functions $\Pi_i^m(f)$. We plot the equilibrium full prices, (f_1^m, f_2^m) on the graph in Figure 3. As the equilibria are all symmetric, with $f_1^m = f_2^m$, each such equilibrium is described by a single point on the dashed line. The solid line in Figure 3 corresponds to the fluid-game equilibrium f^*. Note that the y axis covers only an interval of size 0.3 so that the convergence of f_i^m towards f^* is very quick. For $m = 100$, the distance is less than 0.25 which is, in percentage, less than 4%. The gap for the last point in the graph is less than .4%.

In this model, in which the service providers are modeled as $M/M/1$ queues and the competition is only on full price, the diffusion game is identical to the original game and hence an additional step is not required. Of course, if the arrivals were not Poisson and services non-exponential, the diffusion game and the original game would no longer be identical. Rather, $G/G/1$ approximations would be used to construct a diffusion game that provides approximations for the complex original game.

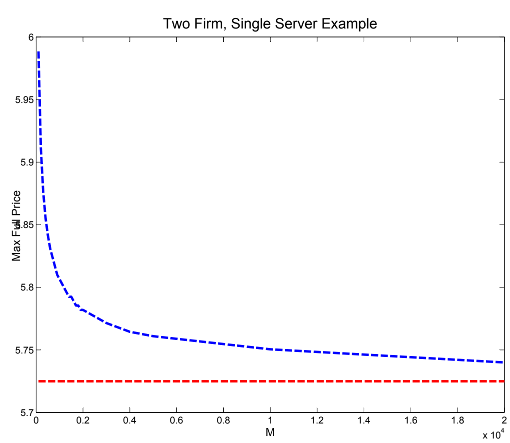

Figure 3: A single server model

Another example of a setting in which our framework is readily applicable is the setting with segmented markets (as considered in Allon and Federgruen (2008)) in which each service provider serves multiple customer classes. In this model, we expect that the available approximations for multi-class queueing systems would be used in the construction of the diffusion game.

Yet another model that seems amenable to analysis through our framework (provided that the market in consideration is large) is one in which the competition is incorporated with learning. These are markets in which the demand characteristics as well as the price and service level actions of all the players in the market are not necessarily observable. Large-scale approximations have been recently used in the context of learning and pricing in revenue management (see e.g. Besbes and Zeevi (2008)) and it seems that these can be combined within our framework to characterize the equilibria in these very realistic, but highly intractable, settings.

Finally, in this paper we considered only the case of linear capacity costs. We made this choice so as not to distract the attention from the main ideas in the proposed framework. Given our framework, the ability to address more general-cost function, as in Borst et al. (2004), would follow from the ability to do so in a monopolist setting. Hence, the fact that such cost functions are treated in the literature on monopolists suggests that these could also be treated in the competitive setting.

Acknowledgments

We are grateful to the reviewers and the AE for their numerous comments which lead to significant improvements over the original manuscript. We also thank Robert Shumsky for helpful discussions in the early stages of this work.

References

Akşin, Z., M. Armony, V. Mehrotra. 2007. The modern call-center: A multi-disciplinary perspective. *Production and Operations Management* **16**(6) 655–688.

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

Allon, G., A. Federgruen. 2008. Competition in service industries with segmented markets. *Management Science, forthcoming* .

Anton, J. 2001. Call center performance benchmark report. Tech. rep., Purdue University–Center for Customer-Driven Quality.

Armony, M., C. Maglaras. 2004. On customer contact centers with a call-back option: Customer decisions, routing rules and system design. *Oper. Res.* **52**(2) 271–292.

Besbes, O., A. Zeevi. 2008. Dynamic pricing without knowing the demand function: Risk bounds and near optimal pricing algorithms. *Oper. Res., forthcoming* .

Borst, S., A. Mandelbaum, M. Reiman. 2004. Dimensioning large call centers. *Oper. Res.* **52**(1) 17–34.

Cachon, G., S. Netessine. 2004. Game theory in supply chain analysis. D. Simchi-Levi, S. D. Wu, Z. J Shen, eds., *Handbook of Quantitative Supply Chain Analysis: Modeling in the E-Business Era*. Kluwer, 13–59.

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

Dasci, A. 2003. Dynamic pricing of perishable assets under competition: a two-period model. *Working Paper. York University, Toronto, Canada* .

Dixon, H.D. 1987. Approximate bertrand equilibria in a replicated industry. *Review of Economoic Studies* **54**(1) 47–62.

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

Gould, N., D. Orban, A. Sartenaer, P.L. Toint. 2002. Componentwise fast convergence in the solution of full-rank systems of nonlinear equations. *Mathematical Programming* **92**(3) 481–508.

Halfin, S., W. Whitt. 1981. Heavy-traffic limits for queues with many exponential servers. *Oper. Res.* **29**(3) 567–587.

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

Jennings, O., F. de Véricourt. 2008. Dimensioning large-scale membership services **56**(1) 173–187.

Kumar, S., R. S. Randhawa. 2008. How heavy should traffic be? revenue optimization with nonlinear delay costs. Working paper, Stanford University and University of Austin.

Levhari, D., I. Luski. 1978. Duopoly pricing and waiting lines. *European Economic Review* **11** 17–35.

Lu, L. X., J. A. Van Mieghem, R. C. Savaskan. 2007. Incentives for quality through endogenous routing. *Manufacturing & Service Operations Management, forthcoming* .

Maglaras, C., A. Zeevi. 2003. Pricing and capacity sizing for systems with shared resources: Scaling relations and approximate solutions. *Management Science* **49**(8) 1018–1038.

Maglaras, C., A. Zeevi. 2005. Pricing and design of differentiated services: Approximate analysis and structural insights. *Oper. Res.* **53**(2) 242–262.

Mandelbaum, A., G. Yom-Tov. 2009. Capacity management in hospitals: Semi-open queueing networks in the qed regime. Working paper, Technion–Israel institute of technology.

Tijs, S.H. 1981. Nash equilibria for noncooperative n-person games in normal form. *SIAM Review* **23**(2) 225–237.

Vives, X. 2000. *Oligopoly pricing: Old ideas and new tools*. The MIT Press, Cambridge, Massachusets.

Whitt, W. 1980. Representation and approximation of noncooperative sequential games. *Siam J. of Control and Optimization* .

Whitt, W. 2003. How multiserver queues scale with growing congestion-dependent demand. *Oper. Res.* **51**(4) 531–542.

Wolff, R.W. 1989. *Stochastic Modeling and the Theory of Queues*. Prentice Hall, Englewood Cliffs, NJ.

Appendix

This appendix is composed of three sections. §A adds some more numerical experiments on top of those conducted in §7 of the paper. §B partially generalizes the results of the paper by removing the

"Diagonal Dominance" condition (C3). Finally, §C is concerned with a global asymptotic stability property of the sequence of market games.

A. Additional numerical examples

This section has two purposes. First, we numerically show that, as the sensitivities to delay increase, the pricing decision of the firms under the diffusion game is close to the fluid game price p^*. Then, we provide another illustration of the quality of the approximation, on top of the one in §7.

First, then, we claim that for fixed capacity costs, increasing the sensitivity of the customers to waiting time, renders it optimal for the firms to provide very high service levels (very short response times), so that the portion $\Lambda_i(p,T)(p_i - c_i - \gamma_i/\mu_i)$ of the firm i's payoff would be close to the fluid game profit $\lambda_i(p,0)(p_i - c_i - \gamma_i/\mu_i)$. Table A provides numerical evidence for this intuition. We use the linear demand model from §7. The table then displays the service levels and prices in the equilibrium of the diffusion game, as a function of the service-level sensitivities κ_i and κ_{ij}. For the large values of these parameters, the diffusion game service levels are 0 (or very close to 0) and the equilibrium prices are equal to the price equilibrium of the fluid game which we also independently calculated using the fluid game.

SL sensitivities	Diffusion-game equilibrium SL	Diffusion-game equilibrium price
$\kappa_i = 0.1$, $\kappa_{ij} = 0.01$	(1.0044,1.0044,1.1903)	(59.2561,59.2561,62.8919)
$\kappa_i = 1$, $\kappa_{ij} = 0.1$	(0.2484,0.2484,0.3144)	(59.2441,59.2441,62.8783)
$\kappa_i = 10$, $\kappa_{ij} = 1$	(0.1429,0.1429,0.1937)	(59.2408,59.2408,62.8736)
$\kappa_i = 20$, $\kappa_{ij} = 2$	(0.0947,0.0947,0.1392)	(59.2406,59.2406,62.8719)
$\kappa_i = 30$, $\kappa_{ij} = 3$	(0.0651,0.0651,0.1062)	(59.2421,59.2421,62.8720)
$\kappa_i = 40$, $\kappa_{ij} = 4$	(0.0441,0.0441,0.0832)	(59.2448,59.2448,62.8733)
$\kappa_i = 50$, $\kappa_{ij} = 5$	(0.0281,0.0281,0.0658)	(59.2485,59.2485,62.8754)
$\kappa_i = 60$, $\kappa_{ij} = 6$	(0.0151 ,0.0151,0.0521)	(59.2529,59.2529,62.8783)
$\kappa_i = 70$, $\kappa_{ij} = 7$	(0.0042,0.0042,0.0407)	(59.2581,59.2581,82.7719)
$\kappa_i = 80$, $\kappa_{ij} = 8$	(0,0,0.0311)	(59.2608,59.2608,62.8850)
$\kappa_i = 90$, $\kappa_{ij} = 9$	(0,0,0.0228)	(59.2614,59.2614,62.8879)
$\kappa_i = 100$, $\kappa_{ij} = 10$	(0,0,0.0154)	(59.2621,59.2621,62.8911)
$\kappa_i = 110$, $\kappa_{ij} = 11$	(0,0,0.0089)	(59.2628,59.2628,62.8946)
$\kappa_i = 120$, $\kappa_{ij} = 12$	(0,0,0.0029)	(59.2636,59.2636,62.8983)
$\kappa_i = 130$, $\kappa_{ij} = 13$	(0,0,0)	(59.2641,59,2641,62.9004)

Table 3: Diffusion-game equilibria as a function of the delay sensitivities

When customers are extremely sensitive to service levels, it is then hard to distinguish between the fluid game and the diffusion game so that the benefits of the diffusion game are less evident. In contrast, when customers have lower sensitivities (but still strictly positive), the tradeoff between capacity costs and loss of market share due to waiting is a meaningful one.

We now provide another numerical example of the linear demand model on top of the one provided in §7. In that example, following the discussion above, we purposefully choose service-level sensitivities that are low. We again consider an industry with $I = 3$ firms, industry benchmark of $\bar{T} = 1$, and cost parameters $c_1 = c_2 = 20, c_3 = 5$, and $\gamma_1 = \gamma_2 = 35, \gamma_3 = 50$. We also assume that $b_i = 10$ for all $i = 1, 2, 3$ and $l_{ij} = 4.75, \forall i \neq j$. Finally, we let $a_1 = a_2 = 205$, $a_3 = 295$, $\kappa_i = 0.1$ for all $i = 1, 2, 3$ and $\kappa_{ij} = 0.01$ for all $i \neq j$.

Figure 4 is the analogue of Figure 2 for the new parameters. It displays the maximal profitable deviations for firm 1. We again observe that profitable deviations from the Nash equilibrium of the diffusion game (p^Λ, T^Λ) are small for both the service-level and the price choices. While for small values of Λ, the deviations are not small, they quickly become negligible as Λ grows.

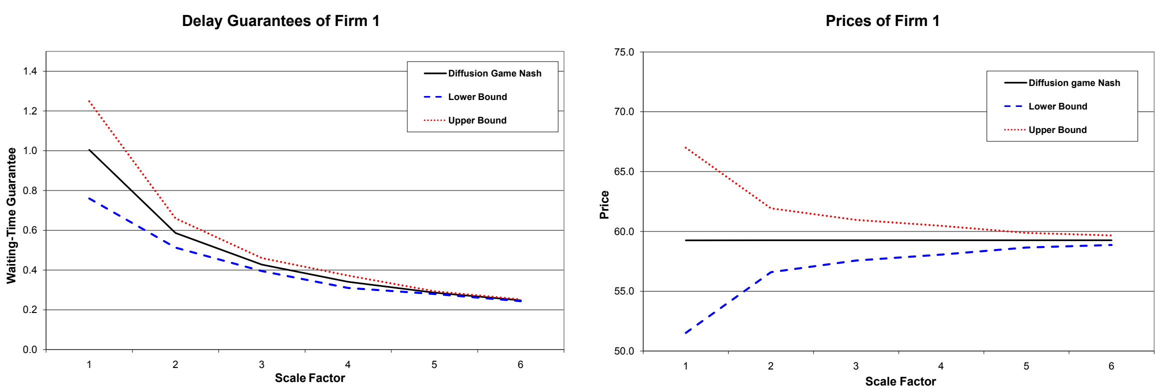

Figure 4: The quality of the approximation – an example of a linear model

B. General continuity of the fluid game

This appendix is concerned with relaxing the "Diagonal Dominance" condition (C3) that is imposed in some of the results in the paper. Motivated by Lemma 5.8, we referred to this property as "linear continuity" of the fluid game. Here we introduce a general notion of continuity for the fluid game together with the corresponding extensions of the results in §5 and §6. The case of linear continuity is then obtained as a special case of the more general setting.

Prior to Theorem 6.7 we introduced the notion of "the fluid game on T". We now assign a number to that definition as will be repeatedly referring to it.

Definition B.1 *The fluid game on T is the I-player game with profit functions*

$$\bar{\Pi}_i^{T,P}(p) := \lambda_i(p,T)\left(p_i - c_i - \frac{\gamma_i}{\mu_i}\right),\ i \in \mathcal{I},$$

and strategy space $\mathcal{P}$.

The fluid game is then the special case with $T = 0$. We let $\psi_i^T(p_{-i})$ be the best response of player i (in the fluid game on T) to prices p_{-i} of the competitors and set

$$\psi^T(p) := (\psi_1^T(p_{-1}), \ldots, \psi_I^T(p_{-I})).$$

If the fluid game on T has a unique Nash equilibrium $p^*(T)$, then it is given by the unique solution to the system of equation $p - \psi^T(p) = 0$. We now have the following definition:

Definition B.2 (**uniform g-continuity around** $T = 0$) *We say that the family of fluid games is uniformly g-continuous around $T = 0$ if there exists a continuous function $g : \mathbb{R}_+^I \to \mathbb{R}_+^I$ with $g(0) = 0$ as well as a constant $\delta > 0$ such that for each $T \in [0, \delta]^I$:* (a) *the fluid game on T has a unique equilibrium $p^*(T)$, and* (b) *fixing $\epsilon \in \mathbb{R}_+^I$,*

$$|p - \psi^T(p)| \leq \epsilon$$

implies that

$$|p - p^*(T)| \leq g(\epsilon).$$

Note that g-continuity of the fluid game is the special case with $T = 0$. Within this general terminology, Lemma 5.8 can be re-stated as saying that, under Condition (C3), the fluid game is g-continuous with the function $g(x) := B^{-1}x/(1 - C)$. We recall that B is the $I \times I$ matrix given by $B_{ii} = 0,\ i = 1, \ldots, I$ and $B_{ij} = 0$ for all $i \neq j$.

Lemma B.3 *Suppose that Assumptions 3.1 and 5.1 hold in addition to conditions* (C1)-(C3). *Then, the fluid game is uniformly g-continuous around $T = 0$ with $g(x) = \frac{1}{1-C}|B^{-1}x|$ and C is the constant from* (C3).

In contrast to the case of linear continuity, for which we were able to show in Lemma 5.8 that condition (C3) is sufficient for the continuity property to hold, the general setting requires the user to identify the corresponding function $g(\cdot)$. Once a function $g(\cdot)$ that satisfies Definition B.2 is identified, the following generalization of Theorem 5.9 can be used. We note that the uniform g-continuity is not required for this result, only the g-continuity of the fluid game on $T = 0$.

Theorem B.4 (**bounds on the distance from the fluid game**) *Suppose that Assumptions 3.1, 4.1 and 5.1 hold. Then, there exists a sequence $\epsilon^\Lambda = O(1/r_1^\Lambda, \ldots, 1/r_I^\Lambda)$ such that, for each Λ, $(p^*, 0)$ is an ϵ^Λ Nash equilibrium for the Λ^{th} market game. Moreover,*

$$T_i^{*,\Lambda}(p^*, 0) \sim r_i^\Lambda,\ i \in \mathcal{I} : \alpha_i > 1, \tag{47}$$

$$T_i^{*,\Lambda}(p^*, 0) = O(r_i^\Lambda),\ i \in \mathcal{I} : \alpha_i = 1, \tag{48}$$

and

$$T_i^{*,\Lambda}(p^*, 0) = o(r_i^\Lambda),\ i \in \mathcal{I} : \alpha_i < 1. \tag{49}$$

If, in addition, the fluid game is g-continuous, then

$$|p_i^{*,\Lambda}(p^*, 0) - p_i^*| = O(g_i(\sqrt{\delta^\Lambda})),\ i \in \mathcal{I}. \tag{50}$$

with $\delta_i^\Lambda = M\left(\frac{1}{\Lambda r_i^\Lambda} + (r_i^\Lambda)^{\alpha_i}\right)$ for some constant $M > 0$.

With $T_i^{*,\Lambda}(p^*, 0)$ replaced everywhere by $T_i^{*,\Lambda}(p^\Lambda, T^\Lambda)$, equations (47)-(49) *hold for any sequence $\{(p^\Lambda, T^\Lambda), \Lambda \geq 0\}$ that satisfy $(p^\Lambda, T^\Lambda) \in \mathcal{P} \times \Theta$ for each Λ.*

The following is the general version of Theorem 6.3 in which we replace condition (C3) with general g-continuity.

Theorem B.5 (**bounds on the distance from the diffusion game**) *Suppose that Assumptions 3.1 and 5.1 hold in addition to conditions* (C1)-(C2) *and* (C4)-(C5) *and let $\{(p^\Lambda, T^\Lambda),\ \Lambda \geq 0\}$ be such that (p^Λ, T^Λ) is a Nash equilibrium for the Λ^{th} diffusion game. If in addition, the fluid game is uniformly g-continuous around $T = 0$, then*

$$T_i^{*,\Lambda}(p^\Lambda, T^\Lambda) = T_i^\Lambda + o(r_i^\Lambda) \text{ and } p_i^{*,\Lambda}(p^\Lambda, T^\Lambda) = p_i^\Lambda + o(g_i(\sqrt{\delta^\Lambda})),$$

where $\delta_i^\Lambda = (r_i^\Lambda)^{\alpha_i}$.

We use η^* from Lemma 6.5 in the following theorem which is a generalization of Theorem 6.7.

Theorem B.6 *Suppose that Assumptions 3.1 and 5.1 hold in addition to conditions* (C1), (C2) *and* (C5). *Assume, in addition, that (i) for each Λ, a unique equilibrium $p^{\Lambda}(\eta^{\Lambda})$ exists for the fluid game on $\eta^{\Lambda} := (\eta_1^* r_1^{\Lambda}, \ldots, \eta_I^* r_I^{\Lambda})$, and (ii) the fluid game is uniformly g-continuous around $T = 0$. Then, there exists a sequence $\epsilon^{\Lambda} = o\left(1/r_1^{\Lambda}, \ldots, 1/r_I^{\Lambda}\right)$ such that$(p^{\Lambda}(\eta^{\Lambda}), \eta^{\Lambda})$ is an ϵ^{Λ}-Nash equilibrium for the Λ^{th} diffusion game and for the Λ^{th} market game. Moreover,*

$$T_i^{*,\Lambda}(p^{\Lambda}(\eta^{\Lambda}), \eta^{\Lambda}) = \eta_i^{\Lambda} + o(r_i^{\Lambda}), \quad \textit{and} \quad p_i^{*,\Lambda}(p^{\Lambda}(\eta^{\Lambda}), \eta^{\Lambda}) = p_i^{\Lambda}(\eta^{\Lambda}) + o\left(g_i(\sqrt{\zeta^{\Lambda}})\right),$$

where $\zeta_i^{\Lambda} = (r_i^{\Lambda})^{\alpha_i}$.

C. Asymptotic Stability

Our results in §6 show that, if the market game is initialized in the diffusion game Nash equilibrium, there will not be a *single step* unilateral deviation that increases the profit by more than $o(r_i^{\Lambda})$ for firm i. It is plausible, however, that if firms play their best responses sequentially multiple times, the market may drift away from a neighborhood of diffusion game equilibrium. Moreover, if the market is initialized away from the diffusion game Nash equilibrium it is a-priori possible that it will never reach the neighborhood of that approximate game.

In this section we show that this cannot happen provided that a global stability condition is imposed on the fluid game. Specifically, we identify a sufficient condition that guarantees that, for any starting point $(p, T) \in \mathcal{P} \times \Theta$, the market will converge to a neighborhood of the diffusion game equilibrium and this neighborhood is exactly the one characterized in Theorems 6.7 and its general version in Theorem B.6. We will show that our approximations are globally asymptotically stable. For an introduction to stability of non-cooperative games, see e.g. Section 2.6 of Vives (2000).

The global stability that we show is frequently referred to as tatônement stability. Fixing Λ, the Tatônement scheme is as follows: the market is initialized at some arbitrary point $(p^{\Lambda,0}, T^{\Lambda,0}) \in \mathcal{P} \times \Theta$. Firms then play their best responses in a fixed sequence, which we assume without loss of generality to be $(1, \ldots, I$. Letting $i(k)$ be the firm that plays its best response in the k^{th} turn $k = 1, \ldots, \infty$, we put $(p^{\Lambda}_{-i(k)}, T^{\Lambda}_{-i(k)}) = (p^{\Lambda}_{-i(k-1)}, T^{\Lambda}_{-i(k-1)})$ and

$$(p^{\Lambda}_{i(k)}, T^{\Lambda}_{i(k)}) \in \underset{p_i \in \mathcal{P}_i, T_i \in [0, \bar{T}]}{\operatorname{argmax}} \Pi_i^{\Lambda}(p_{-i(k-1)}, p, T_{-i(k-1)}, T_i).$$

When the best response of player $i(k)$ is not unique, we choose the maximal one in the sense of the partial ordering of $\mathbb{R}^2_+$, by which $x \geq y$ for $x, y \in \mathbb{R}^2_+$ if $x_1 > y_1$ or if $x_1 = y_1$ and $x_2 > y_2$.

Put now

$$\mathcal{L}^{\Lambda}(n) := \{(p^{\Lambda,k}, T^{\Lambda,k}),\ k = n, n+1, \dots\}. \tag{51}$$

In words, $\mathcal{L}^{\Lambda}(n)$ is the trajectory of the Λ^{th} market game under the Tatônement scheme, excluding the first n steps. If the Λ^{th} market game has a unique Nash equilibrium $(p^{\Lambda}, T^{\Lambda})$ and is globally stable, we would have that $(p^{\Lambda,k}, T^{\Lambda,k}) \to (p^{\Lambda}, T^{\Lambda})$ as $k \to \infty$. Consequently, for each $\epsilon > 0$, there would exist $n(\epsilon)$ such that for all $n \geq n(\epsilon)$

$$\sup_{(p,T)\in\mathcal{L}^{\Lambda}(n)} (\|p - p^{\Lambda}\| + \|T - T^{\Lambda}\|) \leq \epsilon.$$

We, however, do not assume existence of a Nash equilibrium. In particular, we do not require any form of stability for this game. Rather, we only assume the global stability for the fluid game on T (see Definition B.1) for all T small enough. Specifically, we will require that, the fluid game on T has a Nash equilibrium and that, if the players in the fluid game on T follow a Tatônement scheme as above, then the resulting sequence of price vectors, $\{p^{k,T},\ k \geq 1\}$, satisfies $p^{k,T} \to p^*(T)$ as $k \to \infty$. The limit $p^*(T)$ must be the unique equilibrium of the fluid game on T.

The fluid game on T is a simple game in which each player's strategy space is one-dimensional and the payoff functions, under condition (C1), are twice continuously differentiable. It is known (see e.g. §2.6.2 of Vives (2000) and the references therein) that a sufficient condition for the global stability for such a game is the diagonal dominance condition (C3). When the best response functions have derivatives that are continuous in T, it suffices to impose (C3) on the regular fluid game (see Definition 5.1) for it to hold for the fluid game on T, assuming T is small enough. Such a continuity holds, for example, for the multinomial-logit demand model; see Example 6.1. Consequently, Theorem C.1 holds for the multinomial-logit demand model. Similar argument show that it holds for the linear demand model of §7. In our main result of this section we hence assume that (C3) holds for the fluid game on T for all T small enough, i.e, that there exists $C < 1$ such that

$$\sum_{k\in\mathcal{I}} \left| \frac{\partial}{\partial p_k} \psi_i^T(p_{-i}) \right| \leq C,\ p \in \mathcal{P},\ i \in \mathcal{I}.$$

Theorem C.1 *Suppose that Assumptions 3.1 and 5.1 hold in addition to conditions* (C1)-(C3) *and* (C5)*. Assume that for all $\delta > 0$ small enough and any T with $\|T\| \leq \delta$, the fluid game on T satisfies condition* (C3)*. Then, there exists a sequence $\{n^{\Lambda},\ \Lambda \geq 0\}$, such that*

$$\sup_{(p,T)\in\mathcal{L}^{\Lambda}(n^{\Lambda})} \max_{i\in\mathcal{I}} \left| \frac{T_i^{*,\Lambda}(p,T)}{r_i^{\Lambda}} - \eta_i^* \right| \to 0 \text{ as } \Lambda \to \infty, \tag{52}$$

and

$$\sup_{(p,T)\in\mathcal{L}^\Lambda(n^\Lambda)} \max_{i\in\mathcal{I}} \left| \frac{p_i^{*,\Lambda}(p,T) - p^*(\eta^\Lambda)}{\sqrt{(r_i^\Lambda)^{\alpha_i}}} \right| \to 0 \textit{ as } \Lambda \to \infty, \tag{53}$$

where $\eta^\Lambda := (\eta_1^* r_1^\Lambda, \ldots, \eta_I^* r_I^\Lambda)$*,* η^* *is as in Lemma 6.5 and* $p^*(\eta^\Lambda)$ *is the unique Nash equilibrium of the fluid game on* η^Λ*.*

The term asymptotic global stability should be interpreted according to Theorem C.1. That is, we say that the sequence of market games is asymptotically globally stable if the neighborhood that we identified in Theorem B.6 is eventually reached, regardless of the starting point.

The key step in proving Theorem C.1 is to show asymptotic stability in fluid scale, i.e, we show that, for all Λ large enough, the trajectories of the Λ^{th} market game converge to a neighborhood of $(p^*, 0)$.

Theorem C.2 *Suppose that Assumptions 3.1 and 5.1 hold in addition to conditions* (C1)-(C3) *and* (C5)*. Assume that for all* $\delta > 0$ *small enough and any* T *with* $\|T\| \leq \delta$*, the fluid game on* T *satisfies condition* (C3)*. Then, there exist a sequence* $\{n^\Lambda,\ \Lambda \geq 0\}$ *such that*

$$\sup_{(p,T)\in\mathcal{L}^\Lambda(n^\Lambda)} \|T\| \to 0 \textit{ as } \Lambda \to \infty. \tag{54}$$

and

$$\sup_{(p,T)\in\mathcal{L}^\Lambda(n^\Lambda)} \|p - p^*\| \to 0 \textit{ as } \Lambda \to \infty. \tag{55}$$

The proofs of Theorems C.1 and C.2 are relegated to the e-companion.

e-companion for:

Pricing and Dimensioning Competing Large-Scale Service Providers

In this e-companion we provide proofs for all the theorems and lemmas. The proofs of the different results appear in the their order of appearance in the paper. Accordingly, the proofs of the results in the Appendix appear in the second portion of this e-companion. Two auxiliary lemmas–EC.1 and EC.2–appear at the end of the e-companion. In this e-companion, to simplify the notation, we replace the best response notation $T_i^{*,\Lambda}(p,T)$ and $p_i^{*,\Lambda}(p,T)$ with $T_i^{*,\Lambda}$ and $p_i^{*,\Lambda}$ whenever the price and service-level vectors (p,T) are clear from the context.

Proof of Lemma 4.1: By known $M/M/N$ formulas (see e.g. Chapter 5-9 of Wolff (1989)), we know that $\hat{e}_i(\Lambda, T_i)$ satisfies

$$P\{W_i^\Lambda > T_i^\Lambda\} = P\{W_i^\Lambda > 0\}e^{-\mu_i T_i^\Lambda \hat{e}_i(\Lambda, T_i^\Lambda)} \leq \phi.$$

As, $P\{W_i^\Lambda > 0\} \leq 1$ we have that $\hat{e}_i(\Lambda, T_i^\Lambda) \leq -\ln(\phi)/\mu_i T_i^\Lambda$. Hence, $T_i^\Lambda \hat{e}_i(\Lambda_i, T_i^\Lambda) \leq -\ln(\phi) + o(\ln(\phi))$, and the result of the Lemma follows. ■

Proof of Theorem 4.2: The corollary follows from Theorem B.4. Indeed, let $\{(p^\Lambda, T^\Lambda), \Lambda \geq 0\}$ be any sequence such that $(p^\Lambda, T^\Lambda) \in \mathcal{P} \times \Theta$ for each Λ. Then, by the last part of Theorem B.4 we have that $T_i^\Lambda \sim r_i^\Lambda$ if $1/r_i^\Lambda = o(\sqrt{\Lambda})$ and $T_i^\Lambda = O(r_i^\Lambda)$ if $1/r_i^\Lambda \sim \sqrt{\Lambda}$. By Lemma 4.1 we then have that $\hat{e}_i(\Lambda_i, T_i^\Lambda) \sim 1/T_i^\Lambda$ if $1/r_i^\Lambda = o(\sqrt{\Lambda})$ and $\hat{e}_i(\Lambda_i, T_i^\Lambda) \sim \sqrt{\Lambda}$ if $1/r_i^\Lambda \sim \sqrt{\Lambda}$. The bounds on the delay probabilities are now obtained by taking convergent subsequences and applying Proposition 1 in Halfin and Whitt (1981). ■

Proof of Theorem 5.2: The proof draws on Definition 3.2 of ϵ-Nash equilibria, Assumption 5.1 on the uniqueness of the equilibrium p^* for the fluid game, and the properties of the demand functions as listed in Assumption 3.1.

We fix a sequence T^Λ that satisfies the following three properties:

$$\max_{i \in \mathcal{I}} \sup_{p \in \mathcal{P}} |\Lambda_i(p, T^\Lambda) - \Lambda_i(p, 0)| \leq \epsilon^\Lambda/16, \tag{EC1}$$

$$\sup_{p \in \mathcal{P}} \hat{e}_i(\Lambda_i(p, T^\Lambda), T_i^\Lambda) \leq \epsilon^\Lambda/16\gamma_i, \text{ and} \tag{EC2}$$

$$T^\Lambda \to 0, \text{ as } \Lambda \to \infty. \tag{EC3}$$

Such a sequence exists by the absolute continuity of the demand functions on the compact domain $\mathcal{P} \times \Theta$ and by Lemma 4.1. For (EC2) we are using the assumption that $\epsilon^\Lambda \to \infty$.

To show that (p^*, T^Λ) is an ϵ^Λ-Nash equilibria for the Λ^{th} market game, fix a firm i and a sequence $\{(p_i'^\Lambda, T_i'^\Lambda), \Lambda \geq 0\}$ with $(p_i'^\Lambda, T_i'^\Lambda) \in \mathcal{P}_i \times \Theta$ of prices and service levels for firm i such that $(p_i'^\Lambda, T_i'^\Lambda) \neq (p_i^*, T_i^\Lambda)$. Define

$$(\tilde{p}^\Lambda, \tilde{T}^\Lambda) := (p_i'^\Lambda, T_i'^\Lambda) \uparrow (p^*, T^\Lambda)_{-i}.$$

As $\hat{e}_i(\cdot, \cdot) \geq 0$, we have that

$$\Pi_i^\Lambda(\tilde{p}^\Lambda, \tilde{T}^\Lambda) \leq \Lambda_i(\tilde{p}^\Lambda, \tilde{T}^\Lambda) \left(\tilde{p}_i^\Lambda - c_i - \frac{\gamma_i}{\mu_i} \right)$$

By the choice of T^Λ, we have that

$$\left[\Lambda_i(\tilde{p}^\Lambda, \tilde{T}^\Lambda) \left(\tilde{p}_i^\Lambda - c_i - \frac{\gamma_i}{\mu_i} \right) - \Lambda_i(\tilde{p}^\Lambda, 0) \left(\tilde{p}_i^\Lambda - c_i - \frac{\gamma_i}{\mu_i} \right) \right] \leq \epsilon^\Lambda / 8. \tag{EC4}$$

Indeed, one writes $\lambda_i(\tilde{p}^\Lambda, \tilde{T}^\Lambda) - \lambda_i(\tilde{p}^\Lambda, 0) = \lambda_i(\tilde{p}^\Lambda, T^\Lambda) - \lambda_i(\tilde{p}^\Lambda, 0) - \lambda_i(\tilde{p}^\Lambda, T^\Lambda) + \lambda_i(\tilde{p}^\Lambda, \tilde{T}^\Lambda)$. By (EC1) we then have that $|\Lambda_i(\tilde{p}^\Lambda, T^\Lambda) - \Lambda_i(\tilde{p}^\Lambda, 0)| \leq \epsilon^\Lambda/8$. There are now two cases: if $T_i'^\Lambda \leq T_i^\Lambda$ then we can apply (EC1) once again with T^Λ replaced with $\tilde{T}^\Lambda$. If, on the other hand, $T_i'^\Lambda > T_i^\Lambda$, then the monotonicity of the demand functions is invoked to have that $\Lambda_i(\tilde{p}^\Lambda, \tilde{T}^\Lambda) - \Lambda_i(\tilde{p}^\Lambda, T^\Lambda) \leq 0$.

Note that (EC4) is independent of the actual values of the sequence $\{(p_i'^\Lambda, T_i'^\Lambda),\ \Lambda \geq 0\}$ and depends only on the values of $\{(p^\Lambda, T^\Lambda),\ \Lambda \geq 0\}$. By (EC1) we have that

$$\left| \Lambda_i(p^*, T^\Lambda) \left(p_i^* - c_i - \frac{\gamma_i}{\mu_i} \right) - \Lambda_i(p^*, 0) \left(p_i^* - c_i - \frac{\gamma_i}{\mu_i} \right) \right| \leq \epsilon^\Lambda / 8. \tag{EC5}$$

By the definition of p^* as the Nash equilibrium of the fluid game, we have that

$$\Lambda_i(\tilde{p}^\Lambda, 0) \left(\tilde{p}_i^\Lambda - c_i - \frac{\gamma_i}{\mu_i} \right) \leq \Lambda_i(p^*, 0) \left(p_i^* - c_i - \frac{\gamma_i}{\mu_i} \right). \tag{EC6}$$

Combining (EC4), (EC5) and (EC6) we readily have that,

$$\Lambda_i(\tilde{p}^\Lambda, \tilde{T}^\Lambda) \left(\tilde{p}_i^\Lambda - c_i - \frac{\gamma_i}{\mu_i} \right) \leq \Lambda_i(p^*, 0) \left(p_i^* - c_i - \frac{\gamma_i}{\mu_i} \right) + \frac{\epsilon^\Lambda}{4}.$$

Using this together with equation (EC5), we then have that

$$\Pi_i^\Lambda(\tilde{p}^\Lambda, \tilde{T}^\Lambda) \leq \Lambda_i(p^*, T^\Lambda) \left(p_i^* - c_i - \frac{\gamma_i}{\mu_i} \right) + \frac{\epsilon^\Lambda}{2}.$$

By (EC2) $\hat{e}_i(\Lambda_i, T_i^\Lambda) \le \epsilon^\Lambda/2\gamma_i$ and we conclude that

$$\Pi_i^\Lambda(\tilde{p}^\Lambda, \tilde{T}^\Lambda) \le \Pi_i^\Lambda(p^*, T^\Lambda) + \epsilon^\Lambda,$$

so that, for each Λ, (p^*, T^Λ) is an ϵ^Λ-Nash equilibrium. Finally, note that we can also find a sequence T^Λ of vectors with equal components, i.e, such that $T_1^\Lambda = T_2^\Lambda \ldots = T_I^\Lambda$, that satisfies (EC1)-(EC3) and repeat the whole argument with this sequence to get the second part of the Theorem. ■

Proof of Theorem 5.3: We divide the proof into two parts. We first prove the characterization for the equilibrium service-levels in equation (16). We then proceed to prove the equilibrium-price characterization in equation (17).

Proof of (16): Let $(\tilde{p}^\Lambda, \tilde{T}^\Lambda)$ be the sequence of ϵ^Λ-Nash equilibrium. To reach a contradiction, assume that there is no such sequence δ^Λ for T^Λ. In particular, there exists a firm i such that $\limsup_{\Lambda\to\infty} \tilde{T}_i^\Lambda \ge \delta$, for some $\delta > 0$. We may use the compactness of Θ to choose a subsequence Λ^j such that $\lim_{j\to\infty} \tilde{T}_i^{\Lambda^j} = \tilde{\delta} \ge \delta$.

Define $\bar{T}^\Lambda$ by setting $\bar{T}_i^\Lambda = \zeta/\sqrt{\Lambda}$ for this firm i and some $\zeta > 0$ and by setting $\bar{T}_k^\Lambda = T_k^\Lambda$ for all $k \neq i$. Since $\bar{T}_i^\Lambda \to 0$ as $\Lambda \to \infty$, we can re-choose j large enough so that $\bar{T}_i^{\Lambda^j} \le \tilde{T}_i^\Lambda - \eta$, for some $0 < \eta \le \delta$. Since $\lambda_i(p, T)$ is strictly decreasing in T_i (see Assumption 3.1), there exists $\epsilon > 0$, such that

$$\lambda_i(\tilde{p}^{\Lambda^j}, \bar{T}^{\Lambda^j}) \cdot \left(\tilde{p}_i^{\Lambda^j} - c_i - \frac{\gamma_i}{\mu_i}\right) - \lambda_i(\tilde{p}^{\Lambda^j}, \tilde{T}^{\Lambda^j}) \cdot \left(\tilde{p}_i^{\Lambda^j} - c_i - \frac{\gamma_i}{\mu_i}\right) \ge 4\epsilon.$$

Using the definition of the profit functions we have that

$$\begin{aligned} \Pi_i^{\Lambda^j}(\tilde{p}^{\Lambda^j}, \tilde{T}^{\Lambda^j}) - \Pi_i^{\Lambda^j}(\tilde{p}^{\Lambda^j}, \bar{T}^{\Lambda^j}) \quad \le \quad & \Lambda^j \lambda_i(\tilde{p}^{\Lambda^j}, \tilde{T}^{\Lambda^j}) \left(p_i - c_i - \tfrac{\gamma_i}{\mu_i}\right) \\ & - \left(\Lambda^j \lambda_i(\tilde{p}^{\Lambda^j}, \bar{T}^{\Lambda^j}) \cdot \left(\tilde{p}_i^{\Lambda^j} - c_i - \tfrac{\gamma_i}{\mu_i}\right) - \gamma_i \hat{e}_i(\Lambda_i^j, \zeta/\sqrt{\Lambda})\right) \\ & -4\epsilon\Lambda + \gamma_i \hat{e}_i(\Lambda_i^j, \zeta/\sqrt{\Lambda}). \end{aligned}$$

By Lemma 6.1 we have that $\hat{e}_i(\Lambda_i, \zeta/\sqrt{\Lambda}) \le K\sqrt{\Lambda}$ for all Λ large enough and some $K > 0$. Hence, we can re-choose j, so that

$$\Pi_i^{\Lambda^j}(\tilde{p}^{\Lambda^j}, \tilde{T}^{\Lambda^j}) - \Pi_i^{\Lambda^j}(\tilde{p}^{\Lambda^j}, \bar{T}^{\Lambda^j}) \le -2\epsilon\Lambda^j.$$

Firm i can, hence, improve its profit, $\Pi_i^{\Lambda^j}$, by more than $\epsilon\Lambda$. Since $\epsilon^\Lambda/\Lambda \to 0$, there exists j_0 such that $\epsilon^{\Lambda^j} \le \epsilon\Lambda$ for all $j \ge j_0$. Consequently, for all j large enough, $(\tilde{p}^{\Lambda^j}, \tilde{T}^{\Lambda^j})$ can not be an ϵ^Λ-Nash equilibrium.

Proof of (17): We fix the sequence (p^Λ, T^Λ) of ϵ^Λ-Nash equilibria. To reach a contradiction assume that $\limsup_{\Lambda\to\infty} \|p^\Lambda - p^*\| > 0$. We then say that p^Λ is *asymptotically distinguishable from* p^*. Note that if $\max_{i\in\mathcal{I}} \limsup_{\Lambda\to\infty} T_i^\Lambda > 0$, the result of the theorem trivially follows from (16). Hence, we assume $T_i^\Lambda \to 0$ for all $i \in \mathcal{I}$. We will show that under the assumption that p^Λ is distinguishable from p^*, every limit point p of p^Λ must be an equilibrium point for the fluid game. Such a limit point exists by the compactness of $\mathcal{P}$. Since p^Λ is distinguishable from p^*, this will imply the existence of multiple equilibria for the fluid game, contradicting Assumption 5.1. It remains, hence, only to show that every limit point p is indeed an equilibrium point for the fluid game. Towards that end, fix a limit point p of $\{p^\Lambda,\ \Lambda \geq 0\}$ and the corresponding convergent subsequence $\{p^{\Lambda^k},\ k \geq 0\}$. We claim that p is an ϵ-Nash equilibrium for the fluid game for any $\epsilon > 0$. In turn, it is a Nash equilibrium for this game. Define $\bar{p} := (\bar{p}_i, p_{-i})$, for some price $\bar{p}_i \in [p_i^{min}, p_i^{max}]$ with $\bar{p}_i \neq p_i$. Then, since (p^Λ, T^Λ) is the assumed sequence of ϵ^Λ-Nash equilibria, we have that for all k large enough,

$$\bar{\Pi}_i^{\Lambda^k}(\bar{p}^{\Lambda^k}, T^{\Lambda^k}) \leq \bar{\Pi}_i^{\Lambda^k}(p^{\Lambda^k}, T^{\Lambda^k}) + \epsilon/4,$$

for some $\epsilon > 0$. Observe that by Lemma 6.1, $\hat{e}_i(\Lambda_i, T_i^\Lambda)/\Lambda \to 0$ as $\Lambda \to \infty$. This, together with the continuity of the demand functions, implies that

$$\lim_{\Lambda\to\infty} \sum_{i\in\mathcal{I}} \left| \frac{\Pi_i^\Lambda(p^\Lambda, T^\Lambda)}{\Lambda} - \bar{\Pi}_i^P(p) \right| = 0.$$

In particular,

$$\lim_{k\to\infty} \sum_{i\in\mathcal{I}} \left| \frac{\Pi_i^{\Lambda^k}(p^{\Lambda^k}, T^{\Lambda^k})}{\Lambda} - \bar{\Pi}_i^P(p) \right| = 0.$$

Hence,

$$\frac{\Pi_i^{\Lambda^k}(\bar{p}^{\Lambda^k}, T^{\Lambda^k})}{\Lambda} \to \bar{\Pi}_i^P(\bar{p}) \text{ and } \frac{\Pi_i^{\Lambda^k}(\bar{p}^{\Lambda^k}, T^{\Lambda^k})}{\Lambda} \to \bar{\Pi}_i^P(p), \text{ as } k \to \infty$$

where $\bar{p} = (\bar{p}_i, p_{-i})$, and we have that

$$\bar{\Pi}_i^P(\bar{p}) \leq \bar{\Pi}_i^P(p) + \epsilon.$$

Hence, p is an ϵ-Nash equilibrium for the fluid game. As ϵ was arbitrary, we have that p is a Nash equilibrium for the fluid game. Since $p \neq p^*$ we have a contradiction to the uniqueness of equilibria for the fluid game. ■

Proof of Lemma 5.7: As p_{-i} is fixed, it suffices to prove the result for a one dimensional function. Specifically, fix a twice continuously differentiable function $f(x) : \mathbb{X} \to \mathbb{R}_+$ that is defined on a compact and convex set $\mathbb{X} \subseteq \mathbb{R}_+$ and such that $\frac{\partial^2}{\partial x^2} f(x) < 0$ for all $x \in \mathbb{X}$. Let $x^* = \text{argmax}_{x\in\mathbb{X}} f(x)$. We claim that $|x^* - y| \leq C\sqrt{\epsilon}$ whenever $\max_{x\in\mathbb{X}} f(x) - f(y) \leq \epsilon$ for some constant $\epsilon > 0$. The first part of the Lemma will then follow by setting $f(x) := \bar{\Pi}_i(x, p_{-i})$. To prove our claim for $f(x)$ consider first the case in which x^* lies in the interior of $\mathbb{X}$. In this case x^* solves the first order condition $f'(x) = 0$. Assume that $|x^* - y| > \sqrt{\epsilon}$. Assume that $x^* < y$ (the other case is treated similarly). As the second derivative is strictly negative we have that $f'(x^* + \sqrt{\epsilon}/2) \leq -\delta\sqrt{\epsilon}/2$ with δ as in the condition of the Lemma. In particular, as $f'(x)$ is a decreasing function we have that

$$f(x^*) \geq f(x^* + \sqrt{\epsilon}/2) \geq f(y) + (y - x^* - \sqrt{\epsilon}/2)\delta\sqrt{\epsilon}/2 = f(y) + (y - x^*)\delta\sqrt{\epsilon}/2 + o(\sqrt{\epsilon}).$$

Hence, $f(x^*) - f(y) \leq \epsilon$ implies for all ϵ small enough that $y - x^* \leq C\sqrt{\epsilon}$ for some constant $C > 0$.

To complete the argument, assume that x^* is on the boundary of $\mathbb{X}$. Assume that x^* is the smallest element in $\mathbb{X}$ (the proof is similar for the case in which x^* is the largest element). Then, we have that $f'(x) < 0$ for all $x > x^*$ and in particular $f'(x + \sqrt{\epsilon}/2) \leq -\delta\sqrt{\epsilon}/2$. From here we can apply the same arguments as above. ■

Proof of Lemma 5.8: This lemma is a direct consequence of Lemma B.3. ■

Proof of Theorems 5.6 and 5.9: These theorems are special cases of Theorem B.4. ■

Proof of Lemma 6.1: The first part of the lemma follows directly from Proposition 9.3 in Borst et al. (2004) and from item (ii) in Example 9.4 there. We turn to prove the convexity of the function $\beta_i(\eta)$.

Let $f(x, y) := \mathbf{P}(x)e^{-\mu_i x \cdot y}$. By Lemma B.1 in Borst et al. (2004), the function $\mathbf{P}(\cdot)$ is strictly decreasing convex. Using this property one can easily show that the function $f(x, y)$ is convex and strictly decreasing in x and convex and strictly decreasing in $y \in [0, \bar{T}]$. Also, $\frac{\partial}{\partial x}\frac{\partial}{\partial y} f < 0$.

Note that $\beta_i(\eta)$ is the function that satisfies $f(\beta_i(\eta), \eta) = \phi$. By differentiating once on both sides of this equality we get:

$$\frac{\partial}{\partial x} f(\beta_i(\eta), \eta) \cdot \beta_i'(\eta) + \frac{\partial}{\partial y} f(\beta_i(\eta), \eta) = 0.$$

Using the fact that f is strictly decreasing in y and x we then have that $\beta_i'(\eta) < 0$. Differentiating for the second time we have

$$\frac{\partial^2}{\partial x^2} f(\beta_i(\eta), \eta) \cdot (\beta_i'(\eta))^2 + 2\frac{\partial}{\partial x}\frac{\partial}{\partial y} f(\beta_i(\eta), \eta) \cdot \beta_i'(\eta) + \frac{\partial^2}{\partial y^2} f(\beta_i(\eta), \eta) = -\beta_i''(\eta)\frac{\partial}{\partial x} f.$$

The first element on the left-hand side is positive by the convexity of f in x. The second element is positive by the property $\frac{\partial}{\partial x}\frac{\partial}{\partial y} f < 0$. The last element on the left-hand side is positive by the convexity of f in its y argument. Dividing both sides by $\frac{\partial}{\partial x} f$ and using the fact that f is decreasing in x we then have that $\beta_i''(\eta) > 0$ and the proof is complete. ∎

Proof of Example 6.1: We prove that the ML demand model satisfies condition (C5). To this end, after some basic manipulations we get

$$\frac{\lambda_i(p, T_{-i}, x) - \lambda_i(p, T_{-i}, 0)}{x^{\alpha_i}} = \frac{(v_i(p_i, x) - v_i(p_i, 0))(1 - \lambda_i(p, T_{-i}, 0))}{1 + \sum_{j \neq i} v_j(p_j, T_j) + v_i(p_i, 0)},$$

where, for all $j \in \mathcal{I}$, $v_j(p_j, T_j) := e^{a_j(T_j) - b_j p_j}$ and $a_j(T_j) = a_j - k_j(T_j)^{\alpha_j}$. Using Taylor expansion around $z = 0$ for the function $f(z) := e^{a_i - b_i p_i - k_i z}$ we then have that

$$\frac{\lambda_i(p, T_{-i}, x) - \lambda_i(p, T_{-i}, 0)}{x^{\alpha_i}} = \frac{k_i v_i(p_i, 0) x^{\alpha_i}(1 - \lambda_i(p, T_{-i}, 0))}{1 + \sum_{j \neq i} v_j(p_j, T_j) + v_i(p_i, 0)} + o(1)$$

and, consequently,

$$\frac{\lambda_i(p, T_{-i}, x) - \lambda_i(p, T_{-i}, 0)}{x^{\alpha_i}} \to \frac{k_i v_i(p_i, 0)(1 - \lambda_i(p, T_{-i}, 0))}{1 + \sum_{j \neq i} v_j(p_j, T_j) + v_i(p_i, 0)}.$$

∎

Proof of Theorem 6.3: This theorem is a special case of Theorem B.5. ∎

Proof of Lemma 6.5: Fix $i \in \mathcal{I}$. We divide the proof into two cases: (i) $\alpha_i \geq 1$ and (ii) $\alpha_i < 1$.
<u>case (i)–$\alpha_i \geq 1$:</u> Fix a sequence $\{(p^\Lambda, T_{-i}^\Lambda),\ \Lambda \geq 0\}$ such that $(p^\Lambda, T_{-i}^\Lambda) \in \mathcal{P} \times [0, \bar{T}]^{I-1}$ and $(p^\Lambda, T_{-i}^\Lambda) \to (p^*, 0)$. We claim that (C5) implies, that for every $\eta > 0$,

$$\frac{\Lambda_i(p^\Lambda, T_{-i}^\Lambda, \eta r_i^\Lambda) - \Lambda_i(p^\Lambda, T_{-i}^\Lambda, 0)}{1/r_i^\Lambda} \to \eta^{\alpha_i} f_i(p^*, 0), \tag{EC7}$$

where $f_i(\cdot, \cdot)$ is the function from (C5). Moreover, the monotonicity (as a function of η) of both the pre-limit and limit functions, guarantees that the convergence is not only pointwise (for a given η) but rather on compact sets. We will prove (EC7) at the end of the proof and we turn to prove the main assertion of the lemma for the case $\alpha_i \geq 1$. To this end, let $\{(p^\Lambda, T^\Lambda), \Lambda \geq 0\}$ be any sequence with $(p^\Lambda, T^\Lambda) \to (p^*, 0)$ as $\Lambda \to \infty$. We note that the argument in the beginning of the proof of Theorem B.4 can be repeated to show that the best service-level response is of order $O(r_i^\Lambda)$ for firm i. In particular, $T_i^{*,\Lambda} = O(r_i^\Lambda)$ so that we can find $C > 0$ such that $T_i^{*,\Lambda} \in [0, Cr_i^\Lambda]$ and we can write

$$\begin{aligned} T_i^{*,\Lambda} &= \underset{x \in [0, Cr_i^\Lambda]}{\operatorname{argmax}} \left(\Lambda_i(p^\Lambda, T_{-i}^\Lambda, x) - \Lambda_i(p^\Lambda, T_{-i}^\Lambda, 0)\right) \left(p_i^\Lambda - c_i - \frac{\gamma_i}{\mu_i}\right) \\ &+ \beta_i(\eta r_i^\Lambda \sqrt{R_i(p^*,0)}) \sqrt{R_i(p^*,0)} + o(\beta_i(\eta r_i^\Lambda \sqrt{R_i(p^*,0)}) \sqrt{R_i(p^*,0)}). \end{aligned} \quad \text{(EC8)}$$

By Lemma EC.2, the last term disappears when dividing by $1/r_i^\Lambda$ so that using that lemma and (EC7), we have that

$$\frac{T_i^{*,\Lambda}}{r_i^\Lambda} \to \eta(p^*),$$

where $\eta(p^*)$ is as defined in the Lemma. It remains to prove (EC7). To this end, let

$$\mathcal{T}_\delta := \{(p, T_{-i}) \in \mathcal{P} \times [0, \delta]^{I-1} : \|p - p^*\| \leq \delta \text{ and } \sum_{j \neq i} T_j \leq \delta\}.$$

As $(p^\Lambda, T^\Lambda) \to (p^*, 0)$, we have that

$$\limsup_{\Lambda \to \infty} \frac{\Lambda_i(p^\Lambda, T_{-i}^\Lambda, \eta r_i^\Lambda) - \Lambda_i(p^\Lambda, T_{-i}^\Lambda, 0)}{1/r_i^\Lambda} \leq \sup_{(p, T_{-i}) \in \mathcal{T}_\delta} \lim_{\Lambda \to \infty} \frac{\Lambda_i(p, T_{-i}, \eta r_i^\Lambda) - \Lambda_i(p, T_{-i}, 0)}{1/r_i^\Lambda}.$$

We now address the right hand side. Fixing (p, T_{-i}) and recalling that, for $\alpha_i \geq 1$, $r_i^\Lambda = \Lambda^{-\frac{1}{1+\alpha_i}}$ we have that

$$\frac{\Lambda_i(p, T_{-i}, \eta r_i^\Lambda) - \Lambda_i(p, T_{-i}, 0)}{1/r_i^\Lambda} = \eta^{\alpha_i} \frac{\lambda_i(p, T_{-i}, \eta \Lambda^{-\frac{1}{1+\alpha_i}}) - \lambda_i(p, T_{-i}, 0)}{\eta^{\alpha_i} \Lambda^{-\frac{\alpha_i}{1+\alpha_i}}} \to \eta^{\alpha_i} f_i(p, T_{-i}),$$

where the convergence follows from (C5) and the fact that $1/\Lambda \to 0$. In particular,

$$\limsup_{\Lambda \to \infty} \frac{\Lambda_i(p^\Lambda, T_{-i}^\Lambda, \eta r_i^\Lambda) - \Lambda_i(p^\Lambda, T_{-i}^\Lambda, 0)}{1/r_i^\Lambda} \leq \eta^{\alpha_i} \sup_{(p, T_{-i}) \in \mathcal{T}_\delta} f_i(p, T_{-i}).$$

Since δ is arbitrary and $f_i(\cdot, \cdot)$ is assumed to be continuous we have that

$$\limsup_{\Lambda \to \infty} \frac{\Lambda_i(p^\Lambda, T_{-i}^\Lambda, \eta r_i^\Lambda) - \Lambda_i(p^\Lambda, T_{-i}^\Lambda, 0)}{1/r_i^\Lambda} \leq \eta^{\alpha_i} f_i(p^*, 0).$$

A similar argument can be repeated with $\liminf$ instead of $\limsup$ to conclude that

$$\lim_{\Lambda \to \infty} \frac{\Lambda_i(p^\Lambda, T_{-i}^\Lambda, \eta r_i^\Lambda) - \Lambda_i(p^\Lambda, T_{-i}^\Lambda, 0)}{1/r_i^\Lambda} = \eta^{\alpha_i} f_i(p^*, 0).$$

case (ii)–$\alpha_i < 1$: We write, as before,

$$\liminf_{\Lambda\to\infty} \frac{\Lambda_i(p^\Lambda, T^\Lambda_{-i}, \eta r_i^\Lambda) - \Lambda_i(p^\Lambda, T^\Lambda_{-i}, 0)}{1/r_i^\Lambda} \geq \inf_{(p^\Lambda, T^\Lambda_{-i})\in\mathcal{T}_\delta} \lim_{\Lambda\to\infty} \frac{\Lambda_i(p, T_{-i}, \eta r_i^\Lambda) - \Lambda_i(p, T_{-i}, 0)}{1/r_i^\Lambda}.$$

We recall that $r_i^\Lambda = 1/\sqrt{\Lambda}$ for all $i \in \mathcal{I}$ with $\alpha_i \leq 1$. Fixing $(p, T_{-i}) \in \mathcal{T}_\delta$ and $\eta > 0$ we then write

$$\frac{\Lambda_i(p, T_{-i}, \eta r_i^\Lambda) - \Lambda_i(p, T_{-i}, 0)}{1/r_i^\Lambda} = \sqrt{\Lambda}\left(\frac{\eta}{\sqrt{\Lambda}}\right)^{\alpha_i} \frac{\lambda_i(p, T_{-i}, \eta/\sqrt{\Lambda}) - \lambda_i(p, T_{-i}, 0)}{\left(\frac{\eta}{\sqrt{\Lambda}}\right)^{\alpha_i}} \to \infty.$$

Where the divergence follows from the fact that $\alpha_i < 1$ and from Assumption 4.1. Consequently,

$$\liminf_{\Lambda\to\infty} \frac{\Lambda_i(p^\Lambda, T^\Lambda_{-i}, \eta r_i^\Lambda) - \Lambda_i(p^\Lambda, T^\Lambda_{-i}, 0)}{1/r_i^\Lambda} = \infty$$

for all $\eta > 0$. Recalling that $T_i^{*,\Lambda}$ must satisfy (EC8) we must have that $T_i^\Lambda = o(1/\sqrt{\Lambda})$–otherwise, we can improve the profit by setting the delay target to $T_i^\Lambda = 0$.

We conclude this proof by noting that, if (p^Λ, T^Λ) is a Nash equilibrium for the Λ^{th} diffusion game for each Λ, then the whole argument can be repeated with $T_i^\Lambda(p^\Lambda, T^\Lambda_{-i})$ replacing $T_i^{*,\Lambda}$. Here we use the observation that this sequence of equilibria must satisfy $(p^\Lambda, T^\Lambda) \to (p^*, 0)$ as $\Lambda \to \infty$. In that case

$$\begin{aligned} T_i^\Lambda &= \operatorname*{argmax}_{x\in[0, Cr_i^\Lambda]} \left(\Lambda_i(p^\Lambda, T^\Lambda_{-i}, x) - \Lambda_i(p^\Lambda, T^\Lambda_{-i}, 0)\right)\left(p_i^\Lambda - c_i - \frac{\gamma_i}{\mu_i}\right) \\ &+ \beta_i(\eta r_i^\Lambda \sqrt{R_i(p^*, 0)})\sqrt{R_i(p^*, 0)}, \end{aligned}$$

and the proof can be completed identically as the proof for $T_i^{*,\Lambda}$. ■

Proof of Theorem 6.7: This theorem is obtained as a special case of Theorem B.6 by noting that the uniform linear continuity of the fluid game around $T = 0$ follows from Lemma B.3. ■

Proof of Lemma B.3: Fix $\epsilon \in \mathbb{R}^I_+$. Using the fact that $p^* = \psi(p^*)$, we write $p_i - \psi_i(p_{-i}) = p_i - p_i^* - (\psi_i(p_{-i}) - \psi(p^*_{-i}))$. Using (C3) we have that

$$|\psi_i(p_{-i}) - \psi_i(p^*_{-i})| < C\|p_{-i} - p^*_{-i}\|,$$

for $C < 1$. Consequently, $|p - \psi(p)| \leq \epsilon$, implies that

$$\epsilon_i \geq p_i - p_i^* - (\psi_i(p_{-i}) - \psi(p^*_{-i})) \geq (1 - C)\|p_{-i} - p^*_{-i}\|.$$

Hence, for each $i \in \mathcal{I}$, we have that

$$\|p_{-i} - p^*_{-i}\| \leq \frac{\epsilon_i}{1-C}$$

and, consequently, that

$$|B(p-p^*)| \leq \frac{1}{1-C} \cdot \epsilon,$$

where B is the matrix with elements $B_{ii} = 0$ and $B_{ij} = 1$ for all $j \neq i$. In particular, the fluid game is g-continuous with $g(x) = \frac{1}{1-C}|B^{-1}x|$. The uniform g-continuity now follows from (C1). Indeed, the continuity of the derivatives guarantees that, for all T small enough, there exists $C < 1$ such that

$$\sum_{k\in\mathcal{I}} \left|\frac{\partial}{\partial p_k}\psi_i^T(p_{-i})\right| \leq C,\ p \in \mathcal{P},\ i \in \mathcal{I}.$$

■

Proof of Theorem B.4: The outline of the proof is as follows: we first prove that, given any sequence $\{(p^\Lambda_{-i}, T^\Lambda_{-i}),\ \Lambda \geq 0\}$, $T_i^{*,\Lambda}(p^\Lambda, T^\Lambda)$ must satisfy that $T_i^{*,\Lambda}(p^\Lambda, T^\Lambda) \sim r_i^\Lambda$ if $1/r_i^\Lambda = o(\sqrt{\Lambda})$ and $T_i^{*,\Lambda}(p^\Lambda, T^\Lambda) = O(r_i^\Lambda)$ otherwise. This will, in particular, establish (47). Having established (47) we will turn to show that $(p^*, 0)$ is an ϵ^Λ-Nash equilibrium as claimed and establish the bounds in (50).

We start, then, with the treatment of $T_i^{*,\Lambda}(p^\Lambda, T^\Lambda)$. We divide the proof into two cases: (i) firm i has $\alpha_i \geq 1$ and (ii) $\alpha_i < 1$, where α_i is the exponent in Assumption 4.1. First, consider a firm i with $\alpha_i \geq 1$. It is can be verified directly that, for these values of α_i, $r_i^\Lambda \sim z_i^\Lambda$ and that $\Lambda f_i(z_i^\Lambda) \sim \frac{1}{z_i^\Lambda}$. Furthermore, by the definition of z_i^Λ, we have that $f_i'(z_i^\Lambda) = \frac{1}{\Lambda(z_i^\Lambda)^2}$ and we can use the convexity of the function $f(x) = x^{\alpha_i}$ to write

$$\Lambda f_i(T_i^\Lambda) \geq \Lambda f_i(r_i^\Lambda) + \frac{\Lambda}{(r_i^\Lambda)^2}(T_i^\Lambda - r_i^\Lambda),$$

for any $T_i^\Lambda \geq r_i^\Lambda$. In particular, if $T_i^\Lambda / r_i^\Lambda \to \infty$ then

$$\frac{\Lambda f_i(T_i^\Lambda) - \Lambda f_i(r_i^\Lambda)}{1/r_i^\Lambda} \to \infty. \tag{EC9}$$

Let $(p^\Lambda_{-i}, p_i^{*,\Lambda}, T^\Lambda_{-i}, x) := (p_i^{*,\Lambda}, x) \uparrow (p^\Lambda, T^\Lambda)_{-i}$ and note that, by definition,

$$T_i^{*,\Lambda} = \operatorname*{argmax}_x \Lambda_i(p^\Lambda_{-i}, p_i^{*,\Lambda}, T^\Lambda_{-i}, x)\left(p_i - c_i - \frac{\gamma_i}{\mu_i}\right) - \gamma_i \hat{e}_i(\Lambda_i, x),$$

and, in particular, that

$$T_i^{*,\Lambda} = \operatorname*{argmax}_x \left(\Lambda_i(p^\Lambda_{-i}, p_i^{*,\Lambda}, T^\Lambda_{-i}, x) - \Lambda_i(p^\Lambda, T^\Lambda_{-i}, 0)\right)\left(p_i - c_i - \frac{\gamma_i}{\mu_i}\right) - \gamma_i \hat{e}_i(\Lambda_i, x).$$

Using Assumption 4.1 we have that

$$\left(\Lambda_i(p^\Lambda, T_{-i}^\Lambda, T_i^\Lambda) - \Lambda_i(p^\Lambda, T_{-i}^\Lambda, 0)\right)\left(p_i - c_i - \frac{\gamma_i}{\mu_i}\right) - \gamma_i \hat{e}_i(\Lambda_i, T_i^\Lambda) \sim -C_1 \Lambda (T_i^\Lambda)^{\alpha_i} - C_2 \frac{1}{T_i^\Lambda}, \tag{EC10}$$

for some constants C_1 and C_2. Assume that $T_i^\Lambda / r_i^\Lambda \to \infty$. Then, it follows from (EC9) that

$$\frac{\left(\Lambda_i(p^\Lambda, T_{-i}^\Lambda, T_i^\Lambda) - \Lambda_i(p^\Lambda, T_{-i}^\Lambda, 0)\right)(p_i - c_i - \gamma_i/\mu_i) - \gamma_i \hat{e}_i(\Lambda_i, T_i^\Lambda)}{1/r_i^\Lambda} \to -\infty. \tag{EC11}$$

Using (EC10) we have, however, that

$$\limsup_{\Lambda\to\infty} \frac{\left(\Lambda_i(p^\Lambda, T_{-i}^\Lambda, r_i^\Lambda) - \Lambda_i(p^\Lambda, T_{-i}^\Lambda, 0)\right)(p_i - c_i - \gamma_i/\mu_i) - \gamma_i \hat{e}_i(\Lambda_i, r_i^\Lambda)}{1/r_i^\Lambda} > -\infty. \tag{EC12}$$

which leads to a contradiction to the definition of $T_i^{*,\Lambda}$. Hence, for $i \in \mathcal{I}$ with $\alpha_i \geq 1$ we must have $T_i^{*,\Lambda}(p_{-i}^\Lambda, T_{-i}^\Lambda) = O(r_i^\Lambda)$. To complete the proof, we need to show that $\liminf_{\Lambda\to\infty} T_i^\Lambda / r_i^\Lambda > 0$ for all with $1/r_i^\Lambda = o(\sqrt{\Lambda})$. Assume, to reach a contradiction, that $1/r_i^\Lambda = o(\sqrt{\Lambda})$ but $T_i^\Lambda / r_i^\Lambda \to 0$. Then, as $\hat{e}_i(\Lambda_i, T_i^\Lambda) \sim \max\{1/T_i^\Lambda, \sqrt{\Lambda}\}$ we have that

$$\frac{\hat{e}_1(\Lambda, T_i^\Lambda)}{1/r_i^\Lambda} \to \infty.$$

In particular, as $\Lambda_i(p, T)$ is decreasing in T_i we will again have (EC11) which is a contradiction to the definition of $T_i^{*,\Lambda}$. Indeed, we can always use r_i^Λ to obtain a better result as in (EC12). The proof for $\alpha_i \geq 1$ is hence complete.

We now turn to consider the case $\alpha_i < 1$. First, note that with $T_i^{*,\Lambda} = 0$ we have that

$$\limsup_{\Lambda\to\infty} \frac{\left(\Lambda_i(p^\Lambda, T_{-i}^\Lambda, 0) - \Lambda_i(p^\Lambda, T_{-i}^\Lambda, 0)\right)(p_i - c_i - \gamma_i/\mu_i) - \gamma_i \hat{e}_i(\Lambda_i, 0)}{1/r_i^\Lambda} = \frac{\gamma_i \hat{e}_i(\Lambda_i, 0)}{1/r_i^\Lambda} \geq -\infty.$$

Fix a sequence $\{T_i^\Lambda,\ \Lambda \geq 0\}$ with $\liminf_{\Lambda\to\infty} T_i^\Lambda / r_i^\Lambda > 0$. Then, we can apply Assumption 4.1 to get (EC11) once again. Consequently, we must have that $T_i^{*,\Lambda} = o(r_i^\Lambda)$ and the proof of the first part of the Theorem is complete.

We turn now to show that $(p^*, 0)$ is an ϵ^Λ Nash equilibria as well as that the price bounds in (50) hold. To this end, note that

$$|\Pi_i^\Lambda(p_{-i}^*, x, 0, T_i^{*,\Lambda}) - \Pi_i^\Lambda(p_{-i}^*, x, 0, 0)| \leq C\left(\hat{e}_i(\Lambda_i(p^*, 0), 0) + |\Lambda_i(p_{-i}^*, x, 0, T_i^{*,\Lambda}) - \Lambda_i(p_{-i}^*, x, 0, 0)|\right), \tag{EC13}$$

for some constant $C > 0$. Using the first part of the theorem in conjunction with Assumption 4.1 and Lemma 4.1 we then have that

$$\left|\Pi_i^\Lambda(p_{-i}^*, x, 0, T_i^{*,\Lambda}) - \Lambda_i(p_{-i}^*, x, 0, 0)\left(x - c_i - \frac{\gamma_i}{\mu_i}\right)\right| \leq C \frac{1}{r_i^\Lambda}\ . \tag{EC14}$$

Similarly, we have that

$$\left|\Pi_i^\Lambda(p^*,0,T_i^{*,\Lambda}) - \Lambda_i(p^*,0)\left(p_i^* - c_i - \frac{\gamma_i}{\mu_i}\right)\right| \le C\frac{1}{r_i^\Lambda}\,. \tag{EC15}$$

By the definition of p^* we have that

$$\Lambda_i(p_{-i}^*,x,0,0)\left(x - c_i - \frac{\gamma_i}{\mu_i}\right) \le \Lambda_i(p^*,0)\left(p_i^* - c_i - \frac{\gamma_i}{\mu_i}\right)\,.$$

Consequently, we have that, for any x and y,

$$\Pi_i^\Lambda(p_i^*,x,0,y) \le \Pi_i^\Lambda(p^*,0) + \frac{C}{r_i^\Lambda},$$

for some (re-defined) constant $C > 0$ so that $(p^*,0)$ is the claimed ϵ^Λ Nash equilibrium for the Λ^{th} market game. We proceed to provide the bounds in (50). By (EC14) and (EC15) we have that

$$\left|\Pi_i^\Lambda(p_{-i}^*,x,0,T_i^{*,\Lambda}) - \Lambda_i(p_{-i}^*,x,0,0)\left(x - c_i - \frac{\gamma_i}{\mu_i}\right)\right| \le C\frac{1}{r_i^\Lambda},$$

and, using (EC13) and Assumption 4.1 we then have that

$$|\bar{\Pi}_i^P(p^*) - \bar{\Pi}_i^P(p_{-i}^*,p_i^{*,\Lambda}(p^*,0))| \le C\left(\frac{1}{\Lambda r_i^\Lambda} + (r_i^\Lambda)^{\alpha_i}\right),$$

for a re-defined constant $C > 0$. Equation (50) now follows from the g-continuity of the fluid game and Lemma 5.7. ■

Proof of Theorem B.5: Let $\{(p^\Lambda,T^\Lambda),\ \Lambda \ge 0\}$ be a sequence such that (p^Λ,T^Λ) is a Nash equilibrium for the Λ^{th} diffusion game. The first part of the theorem follows directly from Lemma 6.5 and the comment at the end of the proof of that lemma by which we have that, for $\alpha_i \ge 1$, both

$$\frac{T_i^\Lambda}{r_i^\Lambda} \to \eta(p^*) \text{ and } \frac{T_i^{*,\Lambda}}{r_i^\Lambda} \to \eta(p^*). \tag{EC16}$$

In particular, $T_i^{*,\Lambda} = T_i^\Lambda + o(r_i^\Lambda)$. In addition, if $\alpha_i < 1$ then both T_i^Λ and $T_i^{*,\Lambda}$ are $o(r_i^\Lambda)$.

We can now use this to prove the second part of the theorem. Note that, by definition of Nash equilibrium for the diffusion game we have that

$$\begin{aligned} 0 &\ge \hat{\Pi}_i^\Lambda(p_{-i}^\Lambda,T_{-i}^\Lambda,p_i^{*,\Lambda},T_i^{*,\Lambda}) - \hat{\Pi}_i^\Lambda(p^\Lambda,T^\Lambda) \\ &= \Lambda_i(p_{-i}^\Lambda,T_{-i}^\Lambda,p_i^{*,\Lambda},T_i^{*,\Lambda})(p_i^{*,\Lambda} - c_i - \gamma_i/\mu_i) - \Lambda_i(p^\Lambda,T^\Lambda)(p_i^\Lambda - c_i - \gamma_i/\mu_i) \\ &+ \gamma_i\sqrt{R_i(p^*,0)}(\beta_i(T_i^{*,\Lambda}\sqrt{R_i(p^*,0)}) - \beta_i(T_i^\Lambda\sqrt{R_i(p^*,0)})). \end{aligned} \tag{EC17}$$

Using lemma 6.1 we have that

$$\begin{aligned}\Pi_i^\Lambda(p_{-i}^\Lambda, T_{-i}^\Lambda, p_i^{*,\Lambda}, T_i^{*,\Lambda}) - \Pi_i^\Lambda(p^\Lambda, T^\Lambda) &= \Lambda_i(p_{-i}^\Lambda, T_{-i}^\Lambda, p_i^{*,\Lambda}, T_i^{*,\Lambda})(p_i^{*,\Lambda} - c_i - \gamma_i/\mu_i) \\ &- \Lambda_i(p^\Lambda, T^\Lambda)(p_i^\Lambda - c_i - \gamma_i/\mu_i) \\ &+ \gamma_i\sqrt{R_i(p^*,0)}(\beta_i(T_i^{*,\Lambda}\sqrt{R_i(p^*,0)}) - \beta_i(T_i^\Lambda\sqrt{R_i(p^*,0)})) \\ &+ o\left(\gamma_i\sqrt{R_i(p^*,0)}(\beta_i(T_i^{*,\Lambda}\sqrt{R_i(p^*,0)}) - \beta_i(T_i^\Lambda\sqrt{R_i(p^*,0)}))\right)\end{aligned} \tag{EC18}$$

By (EC16) we have that both $T_i^\Lambda = \eta(p^*)r_i^\Lambda + o(r_i^\Lambda)$ and $T_i^{*,\Lambda} = \eta(p^*)r_i^\Lambda + o(r_i^\Lambda)$. Using Lemma EC.2 we then have that

$$\gamma_i\sqrt{R_i(p^*,0)}(\beta_i(T_i^{*,\Lambda}\sqrt{R_i(p^*,0)}) - \beta_i(T_i^\Lambda\sqrt{R_i(p^*,0)})) = o(1/r_i^\Lambda).$$

Also, we can fix $\delta > 0$ so that for all Λ large enough $(\eta_i - \delta)r_i^\Lambda \le T_i^{*,\Lambda} \le (\eta_i + \delta)r_i^\Lambda$ and then use equation (EC7) in the proof of Lemma 6.5 to have that both

$$\frac{\Lambda_i(p_{-i}^\Lambda, T_{-i}^\Lambda, p_i^{*,\Lambda}, (\eta_i + \delta r_i^\Lambda)) - \Lambda_i(p_{-i}^\Lambda, T_{-i}^\Lambda, p_i^{*,\Lambda}, 0)}{1/r_i^\Lambda} \to (\eta_i^* + \delta)^{\alpha_i} f_i(p^*, 0),$$

and

$$\frac{\Lambda_i(p_{-i}^\Lambda, T_{-i}^\Lambda, p_i^{*,\Lambda}, (\eta_i - \delta r_i^\Lambda)) - \Lambda_i(p_{-i}^\Lambda, T_{-i}^\Lambda, p_i^{*,\Lambda}, 0)}{1/r_i^\Lambda} \to (\eta_i^* - \delta)^{\alpha_i} f_i(p^*, 0).$$

Since δ is arbitrary we have that

$$\frac{\Lambda_i(p_{-i}^\Lambda, T_{-i}^\Lambda, p_i^{*,\Lambda}, T_i^\Lambda) - \Lambda_i(p_{-i}^\Lambda, T_{-i}^\Lambda, p_i^{*,\Lambda}, 0)}{1/r_i^\Lambda} \to (\eta_i^*)^{\alpha_i} f_i(p^*, 0).$$

A similar argument is repeated for $T_i^{*,\Lambda}$ to conclude that

$$\frac{\Lambda_i(p^\Lambda, T_{-i}^\Lambda, T_i^{*,\Lambda}) - \Lambda_i(p^\Lambda, T_{-i}^\Lambda, T_i^\Lambda)}{1/r_i^\Lambda} \to 0.$$

Plugging these back into (EC17) and (EC18) we have that

$$\Lambda_i(p_{-i}^\Lambda, T_{-i}^\Lambda, p_i^{*,\Lambda}, T_i^\Lambda)\left(p_i^{*,\Lambda} - c_i - \frac{\gamma_i}{\mu_i}\right) \le \Lambda_i(p^\Lambda, T^\Lambda)\left(p_i^\Lambda - c_i - \frac{\gamma_i}{\mu_i}\right) + o(1/r_i^\Lambda).$$

Dividing by Λ we then have that

$$\left|\bar{\Pi}_i^{T^\Lambda,P}(p_i^{*,\Lambda}) - \bar{\Pi}_i^{T^\Lambda,P}(p_i^\Lambda)\right| \le C\frac{1}{\Lambda r_i^\Lambda}.$$

Note that, as p_i^Λ is the equilibrium price for firm i in the diffusion game, it must be the best response of firm i to $(p_{-i}^\Lambda, T^\Lambda)$. In particular, it must be the equal to best response, $\psi^{T^\Lambda}(p^\Lambda)$, in the fluid game on T^Λ (see Definition B.1). Using Lemma 5.7 and the assumed g-continuity of the fluid game on T^Λ we then have that $|p_i^{*,\Lambda} - p_i^\Lambda| \le g_i(\sqrt{\delta^\Lambda})$ with $\delta^\Lambda = \frac{C}{\Lambda r_i^\Lambda}$. Recalling that $\Lambda r_i^\Lambda = (r_i^\Lambda)^{\alpha_i}$ (see e.g. the proof of Theorem B.4) we have the result of the theorem. ■

Proof of Theorem B.6: From Lemma 6.5 we have, for any sequence $(p^\Lambda, T^\Lambda) \to (p^*, 0)$, that

$$\frac{T_i^{*,\Lambda}(p^\Lambda, T^\Lambda)}{r_i^\Lambda} \to \eta_i^*,\ i \in \mathcal{I}. \tag{EC19}$$

Let $\eta^\Lambda = (\eta_1^* r_1^\Lambda, \ldots, \eta_I^* r_I^\Lambda)$ where $\{\eta_i^*,\ i \in \mathcal{I}\}$ are the constant from Lemma 6.5. Consider now an arbitrary sequence of prices $\{p^\Lambda,\ \Lambda \geq 0\}$ with $p^\Lambda \to p^*$ and consider the sequence $\{(p^\Lambda, \eta^\Lambda)\ \Lambda \geq 0\}$. Let $T^\Lambda) = T_i^{*,\Lambda}(p^\Lambda, \eta^\Lambda) \uparrow \eta^\Lambda$. Then,

$$\begin{aligned} \hat{\Pi}_i^\Lambda(p^\Lambda, T^\Lambda) - \hat{\Pi}^\Lambda(p^\Lambda, \eta^\Lambda) &= \Lambda_i(p^\Lambda, T^\Lambda)\left(p_i^\Lambda - c_i - \frac{\gamma_i}{\mu_i}\right) - \Lambda_i(p^\Lambda, \eta^\Lambda)\left(p_i^\Lambda - c_i - \frac{\gamma_i}{\mu_i}\right) \\ &+ \gamma_i\sqrt{R_i(p^*, 0)}(\beta_i(T_i^{*,\Lambda}\sqrt{R_i(p^*, 0)}) - \beta_i(\eta_i^\Lambda\sqrt{R_i(p^*, 0)})) \end{aligned} \tag{EC20}$$

By (EC19) we have that $T_i^{*,\Lambda} = \eta_i^\Lambda + o(r_i^\Lambda)$ so that (by Lemma EC.1)

$$\gamma_i\sqrt{R_i(p^*, 0)}(\beta_i(T_i^{*,\Lambda}\sqrt{R_i(p^*, 0)}) - \beta_i(T_i^\Lambda\sqrt{R_i(p^*, 0)})) = o(1/r_i^\Lambda). \tag{EC21}$$

From equation (EC7) in the proof of Lemma 6.5 we have that

$$\frac{\Lambda_i(p^\Lambda, \eta_{-i}^\Lambda, \eta_i^* r_i^\Lambda) - \Lambda_i(p^\Lambda, \eta_{-i}^\Lambda, 0)}{1/r_i^\Lambda} \to (\eta_i^*)^{\alpha_i} f_i(p^*, 0).$$

Since $T_i^{*,\Lambda} = \eta_i^\Lambda + o(r_i^\Lambda)$, we can fix $\delta > 0$ so that for all Λ large enough $(\eta_i - \delta)r_i^\Lambda \leq T_i^{*,\Lambda} \leq (\eta_i + \delta)r_i^\Lambda$. We can then use again (EC7) and the fact that $\delta > 0$ is arbitrary to conclude that

$$\frac{\Lambda_i(p^\Lambda, T_{-i}^\Lambda, T_i^{*,\Lambda}) - \Lambda_i(p^\Lambda, T_{-i}^\Lambda, 0)}{1/r_i^\Lambda} \to (\eta_i^*)^{\alpha_i} f_i(p^*, 0).$$

In particular,

$$\frac{\Lambda_i(p^\Lambda, T^\Lambda) - \Lambda_i(p^\Lambda, \eta^\Lambda)}{1/r_i^\Lambda} \to 0. \tag{EC22}$$

Consequently, for each sequence $\{p^\Lambda,\ \Lambda \geq 0\}$ with $p^\Lambda \to p^*$, we have that

$$\hat{\Pi}_i^\Lambda(p^\Lambda, \eta_{-i}^\Lambda, T_i^{*,\Lambda}) \leq \hat{\Pi}_i^\Lambda(\tilde{p}^\Lambda, \eta^\Lambda) + o(r_i^\Lambda).$$

In particular, fix $\check{T}_i^\Lambda$ and let $\tilde{T}^\Lambda = \check{T}_i^\Lambda \uparrow \eta^\Lambda$. Then,

$$\hat{\Pi}_i^\Lambda(p^\Lambda, \tilde{T}^\Lambda) \leq \hat{\Pi}_i^\Lambda(p^\Lambda, T^\Lambda) \leq \hat{\Pi}_i^\Lambda(p^\Lambda, \eta^\Lambda) + o(r_i^\Lambda), \tag{EC23}$$

where the first inequality follows from the definition of $T_i^{*,\Lambda}$ as the best service-level response and the second inequality follows from (EC22).

Let $p^\Lambda(\eta^\Lambda)$ be an equilibrium of the fluid game on η^Λ. Then, we claim that

$$p^\Lambda(\eta^\Lambda) \to p^*. \tag{EC24}$$

We will show (EC24) momentarily. We first use it to complete the proof of the theorem. Since $p^\Lambda(\eta^\Lambda) \to p^*$, equation (EC23) holds with p^Λ there replaced by $p^\Lambda(\eta^\Lambda)$. Fix now $\check{p}_i^\Lambda \neq p_i^\Lambda(\eta^\Lambda)$ and let $\tilde{p}^\Lambda = \check{p}_i^\Lambda \uparrow p^\Lambda$. Then, using (EC23), we have that

$$\hat{\Pi}_i^\Lambda(\tilde{p}^\Lambda, \tilde{T}^\Lambda) \leq \hat{\Pi}_i^\Lambda(\tilde{p}^\Lambda, \eta^\Lambda) + o(r_i^\Lambda).$$

Now, by the definition of p^Λ as the solution to the pricing game on η^Λ we have that

$$\hat{\Pi}_i^\Lambda(\tilde{p}^\Lambda, \tilde{T}^\Lambda) \leq \hat{\Pi}_i^\Lambda(\tilde{p}^\Lambda, \eta^\Lambda) + o(r_i^\Lambda) \leq \hat{\Pi}_i^\Lambda(p^\Lambda, \eta^\Lambda) + o(r_i^\Lambda),$$

so that $(p^\Lambda, \eta^\Lambda)$ is the claimed ϵ^Λ-Nash equilibrium for the Λ^{th} diffusion game. The bound on the deviation in service levels follows from the (already established) equality $T_i^{*,\Lambda} = \eta_i^\Lambda + o(r_i^\Lambda)$. To establish the bounds on prices, let $p_i^{*,\Lambda}(p^\Lambda, \eta^\Lambda)$ be the best response of firm i. Then, since we already showed that $(p^\Lambda, \eta^\Lambda)$ is an ϵ^Λ Nash equilibrium we have that

$$\begin{aligned} o(r_i^\Lambda) \geq \hat{\Pi}_i^\Lambda(\tilde{p}^\Lambda, \tilde{T}^\Lambda) - \hat{\Pi}^\Lambda(p^\Lambda, \eta^\Lambda) \quad &= \quad \Lambda_i(\tilde{p}^\Lambda, \tilde{T}^\Lambda)\left(\tilde{p}_i^\Lambda - c_i - \frac{\gamma_i}{\mu_i}\right) - \Lambda_i(p^\Lambda, \eta^\Lambda)\left(p_i^\Lambda - c_i - \frac{\gamma_i}{\mu_i}\right) \\ &+ \quad \gamma_i\sqrt{R_i(p^*,0)}(\beta_i(T_i^{*,\Lambda}\sqrt{R_i(p^*,0)}) - \beta_i(\eta_i^\Lambda\sqrt{R_i(p^*,0)})). \end{aligned}$$

Here $\tilde{T}^\Lambda = T_i^{*,\Lambda} \uparrow \eta^\Lambda$ and $\tilde{p}^\Lambda = p_i^{*,\Lambda} \uparrow p^\Lambda$. Using (EC22) with $\tilde{p}^\Lambda$ and using (EC21) we then have that

$$\left|\Lambda_i(\tilde{p}^\Lambda, \eta^\Lambda)\left(\tilde{p}_i^\Lambda - c_i - \frac{\gamma_i}{\mu_i}\right) - \Lambda_i(p^\Lambda, \eta^\Lambda)\left(p_i^\Lambda - c_i - \frac{\gamma_i}{\mu_i}\right)\right| = o(r_i^\Lambda).$$

Dividing by Λ we then have that

$$\left|\bar{\Pi}_i^{\eta^\Lambda,P}(p_i^{*,\Lambda}) - \bar{\Pi}_i^{T^\Lambda,P}(p_i^\Lambda)\right| \leq C\frac{1}{\Lambda r_i^\Lambda},$$

where $\bar{\Pi}_i^{\eta^\Lambda,P}(\cdot)$ is the profit function of the fluid game on η^Λ (see Definition B.1). But, by definition, p^Λ was the unique equilibrium of the fluid game on η^Λ so that we may apply Lemma 5.7 and the g-continuity of the fluid game on η^Λ to conclude that $p_i^{*,\Lambda} = p_i^\Lambda(\eta^\Lambda) + o\left(g_i(\sqrt{\zeta^\Lambda})\right)$ as claimed. Here we used also the fact that $\Lambda r_i^\Lambda = (r_i^\Lambda)^{\alpha_i}$ (see e.g. the proof of Theorem B.4). To establish that $(p^\Lambda(\eta^\Lambda), \eta^\Lambda)$ is also an ϵ^Λ-Nash for the Λ^{th} market game, one repeats the same arguments with the addition of using Lemma 6.1 to relate $\Pi_i^\Lambda(\cdot,\cdot)$ to $\hat{\Pi}_i^\Lambda(\cdot,\cdot)$.

To complete the proof it remains to establish (EC24). The proof of this claim is very similar to the proof of the second part of Theorem 5.3. We provide the detailed argument for completeness. Towards this end, let $p^\Lambda(\eta^\Lambda)$ be an equilibrium for the fluid game on η^Λ and consider a convergent subsequence $\{p^{\Lambda^j}(\eta^{\Lambda^j}),\ j \geq 1\}$ of $\{p^\Lambda(\eta^\Lambda),\ \Lambda \geq 0\}$. Such a sequence exists by the compactness of $\mathcal{P}$. Let $\tilde{p}$ be its limit. By definition, a Nash equilibrium of the fluid game on η^Λ satisfies, for each $\hat{p}_i \in \mathcal{P}_i$, that

$$\bar{\Pi}_i^{\eta^{\Lambda^j},P}(\check{p}^{\Lambda^j}) \leq \bar{\Pi}_i^{\eta^{\Lambda^j},P}(p^{\Lambda^j}), \tag{EC25}$$

where $\check{p}^\Lambda = \hat{p}_i \uparrow p^{\Lambda^j}$. In particular, we can choose $\hat{p}_i \neq \tilde{p}_i$. Since the demand functions are continuous in their arguments and since $\eta^\Lambda \to 0$ we can take limits on both sides of (EC25) to conclude that

$$\bar{\Pi}_i^{0,P}(\hat{p}_i \uparrow \tilde{p}) \leq \bar{\Pi}_i^{0,P}(\tilde{p}).$$

Since we can repeat this for each $i \in \mathcal{I}$, we have that $\tilde{p}$ is an equilibrium for the fluid game on $T = 0$ which is the fluid game from Definition 5.1. But, by assumption, the fluid game has a unique equilibrium p^* so that we must have $\tilde{p} = p^*$. Since the same argument holds for any convergent subsequence of $\{p^\Lambda(\eta^\Lambda),\ \Lambda \geq 0\}$ we have that $p^\Lambda(\eta^\Lambda) \to p^*$ as required. ■

Lemma EC.1 *Let $\{T_i^\Lambda,\ \Lambda \geq 0\}$ be a sequence such that $T_i^\Lambda \to 0$. Then,*

$$\beta_i(T_i^\Lambda\sqrt{R_i(p^*,0)})\sqrt{R_i(p^*,0)} \sim \min\left\{\sqrt{\Lambda}, \frac{1}{T_i^\Lambda}\right\}$$

Proof: By definition,

$$\mathbf{P}(\beta_i(T_i^\Lambda\sqrt{R_i(p^*,0)}))e^{-\mu_i\beta_i(T_i^\Lambda\sqrt{R_i(p^*,0)})T_i^\Lambda\sqrt{R_i(p^*,0)}} = \phi. \tag{EC26}$$

Now, $\beta_i(x)$ is smaller than $\beta_i(0)$ and $\beta_i(0)$ is a strictly positive constant. By Proposition 1 in Halfin and Whitt (1981) we have that $\mathbf{P}(\beta_i(0)) \in (0,1)$ We now have two cases: first, if $T_i^\Lambda\sqrt{R_i(p^*,0)} \to 0$, then

$$\beta_i(T_i^\Lambda\sqrt{R_i(p^*,0)}) \to \beta_i(0),$$

and, in particular,

$$\beta_i(T_i^\Lambda\sqrt{R_i(p^*,0)})\sqrt{R_i(p^*,0)} \sim \sqrt{\Lambda}.$$

If, on the other hand, $\liminf_{\lambda\to\infty} T_i^\Lambda\sqrt{R_i(p^*,0)} > 0$, then we can use the fact that $\mathbf{P}(\beta_i(0))$ is strictly positive–and that, consequently, $\mathbf{P}(\beta_i(x))$ is strictly positive for any $x \geq 0$–together with (EC26) to conclude that

$$\mu_i\beta_i(T_i^\Lambda\sqrt{R_i(p^*,0)})T_i^\Lambda\sqrt{R_i(p^*,0)} \in [C_1, C_2],$$

for two strictly positive constants C_1 and C_2. From here it follows that

$$\beta_i(T_i^\Lambda\sqrt{R_i(p^*,0)})T_i^\Lambda\sqrt{R_i(p^*,0)} \sim 1/T_i^\Lambda.$$

■

Lemma EC.2 *Fix $\alpha_i \geq 1$ and let $r_i^\Lambda = \Lambda^{-\frac{1}{1+\alpha_i}}$. Then,*

$$\frac{\beta_i(\sqrt{R_i(p^*,0)}\eta r_i^\Lambda)\sqrt{R_i(p^*,0)}}{1/r_i^\Lambda} \to \tilde{\beta}_i(\eta)\left(\frac{\lambda_i(p^*,0)}{\mu_i}\right)^{\frac{1}{1+\alpha_i}}$$

uniformly on compacts sets where, given η, $\tilde{\beta}_i(\eta)$ is the unique solution to

$$\mathbf{P}(\tilde{\beta}_i)e^{-\mu_i\tilde{\beta}_i\eta\sqrt{\frac{\lambda_i(p^*,0)}{\mu_i}}} = \phi,$$

if $\alpha_i = 1$ and it is the unique solution of $e^{-\mu_i\tilde{\beta}_i\eta} = \phi$ otherwise.

Proof: First, we prove that the convergence holds pointwise, i.e, for any fixed η. To this end, assume first that $\alpha_i = 1$. In this case we have that $\eta r_i^\Lambda\sqrt{R_i(p^*,0)} = \eta\sqrt{\lambda_i(p^*,0)}/\mu_i$, and the result of the lemma for this case follows from the definition of $\beta_i(\cdot)$.

Assume now that $\alpha_i > 1$. Then, by Lemma EC.1, we have that $\beta_i(\eta r_i^\Lambda\sqrt{R_i(p^*,0)})\sqrt{R_i(p^*,0)} \sim 1/r_i^\Lambda$. In particular,

$$\beta_i(\eta r_i^\Lambda\sqrt{R_i(p^*,0)} \sim \frac{1}{\sqrt{\Lambda}r_i^\Lambda} \to 0.$$

In particular,

$$\mathbf{P}(\beta_i(\eta r_i^\Lambda\sqrt{R_i(p^*,0)})) \to 1,$$

by Proposition 1 in Halfin and Whitt (1981). Using the definition of $\beta_i(\cdot)$ we have that

$$-\mu_i\beta_i\eta r_i^\Lambda\sqrt{R_i(p^*,0)} = \ln(\phi) + o(1),$$

so that

$$\beta_i\sqrt{R_i(p^*,0)} = -\frac{\ln(\phi)+o(1)}{\mu_i\eta r_i^\Lambda}$$

and the result now follows by dividing by $1/r_i^\Lambda$. ■

We now turn to the proofs of Theorem C.1 and C.2. Since the former uses some ideas of the latter, we prove them in reverse order.

Proof of Theorem C.2: The proof is based on a coupling argument. We couple the market game with the fluid game and apply the assumed global stability of the fluid game to show that, for each Λ large enough, the Λ^{th} market game itself converges to some neighborhood of $(p^*,0)$ under the Tatônement scheme.

To this end, fix $\delta > 0$ and note that by Theorem 4.2,

$$\sup_{(p,T)\in\mathcal{P}\times\Theta} T_i^{*,\Lambda}(p,T) \leq \delta, \tag{EC27}$$

for all Λ large enough. Hence, (54) follows immediately. Also note that, by Lemma 4.1,

$$\sup_{(p,T)\in\mathcal{P}\times\Theta} \hat{e}_i(\Lambda_i(p,T),T_i) \leq \delta\Lambda, \tag{EC28}$$

for all Λ large enough. We define $\Lambda(\delta)$ to be such that both (EC27) and (EC28) hold for all $\Lambda \geq \Lambda(\delta)$. Also, using the assumed continuity of the demand functions, we then have that for all $(p,T) \in \mathcal{P} \times \Theta$ such that $\|T\| \leq \delta$,

$$\bar{\Pi}_i^P(p) - \epsilon(\delta) \leq \frac{\Pi_i^\Lambda(p,T)}{\Lambda} \leq \bar{\Pi}_i^P(p) + \epsilon(\delta), \tag{EC29}$$

where $\epsilon(\delta) \to 0$ as $\delta \to 0$.

Fix now a point $(p^0, T^0) \in \mathcal{P}\times\Theta$. Let $\{\bar{p}^k, k \geq 1\}$ be the trajectory of the fluid game generated by a Tatônement scheme initialized at the point p^0. Let $\{(p^{k,\Lambda}, T^{k,\Lambda}), k \geq 1\}$ be the trajectory of the market game under the corresponding Tatônement scheme initialized at (p^0, T^0). Then, we have the following lemma.

Lemma EC.3 *Given $\delta > 0$ there exists $\Lambda(\delta)$ and $\epsilon(\delta)$ such that*

$$\max_{i\in\mathcal{I}} \left| \frac{\Pi_i^\Lambda(p^{k,\Lambda}, T^{k,\Lambda})}{\Lambda} - \bar{\Pi}_i^P(\bar{p}^k) \right| \leq \epsilon(\delta), \tag{EC30}$$

and

$$\max_{i\in\mathcal{I}} |\bar{p}_i^k - p_i^{k,\Lambda}| \leq M\sqrt{\epsilon(\delta)}, \tag{EC31}$$

for all $k \geq 1$ and $\Lambda \geq \Lambda(\delta)$, and where $M > 0$ is the constant from Lemma 5.7. Moreover, $\epsilon(\delta) \to 0$ as $\delta \to 0$.

We postpone the proof of this lemma to the end of this section and apply it now to complete the proof of the theorem.

Using the assumed global stability of the fluid game we have that $\bar{p}^k \to p^*$ as $k \to \infty$ where p^* is the unique equilibrium of the fluid game. Consequently, there exists $k(\delta)$ large enough so that $\|\bar{p}^k - p^*\| \leq M\sqrt{\epsilon(\delta)}$. Using (EC31) when then have that for all $\Lambda \geq \Lambda(\delta)$ and all $k \geq k(\delta)$, $\|p^{k,\Lambda} - p^*\| \leq IM\sqrt{\epsilon(\delta)}$. Since the same holds for each $\delta > 0$, we can find a sequence ϱ^Λ such that $\varrho^\Lambda \to 0$ as $\Lambda \to \infty$ and so that $\|p^{k,\Lambda} - p^*\| \leq IM\sqrt{k\epsilon(\varrho^\Lambda)}$, for all $k \geq \Lambda(\varrho^\Lambda)$ and $k \geq n^\Lambda := k(\varrho^\Lambda)$. In particular,

$$\sup_{(p,T)\in\mathcal{L}^\Lambda(n^\Lambda)} \|p - p^*\| \to 0 \text{ as } \Lambda \to \infty.$$

This concludes the proof of the theorem. ■

Proof of Theorem C.1: Equation (52) is a direct consequence of Lemma 6.5 and Theorem C.2. Indeed, by Theorem C.2 guarantees, for each Λ large enough, there exists k^{Λ}, such that $\|p^{k,\Lambda} - p^*| + \|T^{k,\Lambda}| \leq \delta$ for all $k \geq k^{\Lambda}$. Lemma 6.5 then guarantees that $T_i^{*,\Lambda}(p^{\Lambda}, T^{\Lambda})/r_i^{\Lambda} \to \eta_i^*$ as $\Lambda \to \infty$ for any sequence $(p^{\Lambda}, T^{\Lambda})$ such that $(p^{\Lambda}, T^{\Lambda}) \to (p^*, 0)$ as $\Lambda \to \infty$. The proof of Lemma 6.5 reveals that the result can be easily strengthened to argue that: given $\epsilon > 0$, there exists $\delta > 0$ such that for all Λ large enough, $|T_i^{*,\Lambda}(p,T)/r_i^{\Lambda} - \eta_i^*| \leq \epsilon$ whenever $|p - p^*| + \|T\| \leq \delta$. This, in turn, implies (52) and we turn to prove (53).

To this end, we use a coupling argument as in the proof of Theorem C.2. We start by fixing $\delta > 0$ and noting that, by (52) and Theorem C.2, there exist a sequence $\{n^{\Lambda},\ \Lambda \geq 0\}$ such that for all Λ large enough and all $(p,T) \in \mathcal{L}^{\Lambda}(n^{\Lambda})$,

$$T_i \in [\eta_i^{\Lambda} - \delta r_i^{\Lambda}, \eta_i^{\Lambda} + \delta r_i^{\Lambda}], \tag{EC32}$$

where we put $\eta^{\Lambda} := (\eta_1^* r_1^{\Lambda}, \ldots, \eta_I^* r_I^{\Lambda})$ and η^* is as in Lemma 6.5. Using Lemma 6.1, we then have that

$$\begin{aligned}
&\beta_i(\sqrt{R_i(p^*,0)}(\eta_i^{\Lambda} + \delta r_i^{\Lambda})\sqrt{R_i(p^*,0)} - \delta/r_i^{\Lambda} \\
&\quad \leq \hat{e}_i(\Lambda_i(p,T), T_i) \\
&\qquad \leq \beta_i(\sqrt{R_i(p^*,0)}(\eta_i^{\Lambda} - \delta r_i^{\Lambda})\sqrt{R_i(p^*,0)} + \delta/r_i^{\Lambda}.,
\end{aligned} \tag{EC33}$$

for all $(p,T) \in \mathcal{L}^{\Lambda}(n^{\Lambda})$ and all Λ large enough.

Let $\Lambda(\delta)$ be large enough so that both (EC32) and (EC33) for all $\Lambda \geq \Lambda(\delta)$ and $n \geq n^{\Lambda}$. Using condition (C5) and (52) we have that there exists a constant $\tau > 0$ such that

$$\begin{aligned}
\Lambda_i(p, \eta_i^{\Lambda})\left(p_i - c_i - \frac{\gamma_i}{\mu_i}\right) - \tau\Lambda\tilde{\epsilon}(\delta)(r_i^{\Lambda})^{\alpha_i} \quad &\leq \quad \Lambda_i(p,T)\left(p_i - c_i - \frac{\gamma_i}{\mu_i}\right) \\
&\leq \quad \Lambda_i(p, \eta_i^{\Lambda})\left(p_i - c_i - \frac{\gamma_i}{\mu_i}\right) + \tau\Lambda\tilde{\epsilon}(\delta)(r_i^{\Lambda})^{\alpha_i},
\end{aligned}$$

where $\tilde{\epsilon}(\delta) \to 0$ as $\delta \to 0$. Consequently,

$$\begin{aligned}
&\Lambda\bar{\Pi}_i^{\eta^{\Lambda},P}(p) + \gamma_i\beta_i(\sqrt{R_i(p^*,0)}\eta_i^{\Lambda} + \delta r_i^{\Lambda})\sqrt{R_i(p^*,0)} - \delta/r_i^{\Lambda} - c\Lambda\tilde{\epsilon}(\delta)(r_i^{\Lambda})^{\alpha_i} \\
&\quad \leq \Pi_i^{\Lambda}(p,T) \leq \\
&\qquad \Lambda\bar{\Pi}_i^{\eta^{\Lambda},P}(p) - \gamma_i\beta_i(\sqrt{R_i(p^*,0)}\eta_i^{\Lambda} - \delta r_i^{\Lambda})\sqrt{R_i(p^*,0)} + \delta/r_i^{\Lambda} + c\Lambda\tilde{\epsilon}(\delta)(r_i^{\Lambda})^{\alpha_i},
\end{aligned}$$

for each $(p,T) \in \mathcal{L}^{\Lambda}(n^{\Lambda})$

Now, by Lemma 6.1, the function $\beta_i(\cdot)$ is continuous so that there exists $\epsilon(\delta)$ (with $\epsilon(\delta) \to 0$ as $\delta \to 0$) and such that

$$\sup_{x\in[-\delta,\delta]} \left|\beta_i(\sqrt{R_i(p^*,0)}\eta_i^\Lambda - xr_i^\Lambda)\sqrt{R_i(p^*,0)} - \beta_i(\sqrt{R_i(p^*,0)}\eta_i^\Lambda)\sqrt{R_i(p^*,0)}\right| \le \epsilon(\delta)/r_i^\Lambda.$$

By the definition of r_i^Λ in (9) we have that $\Lambda r_i^\Lambda = (r_i^\Lambda)^{\alpha_i}$ for all $\alpha_i \ge 1$ and $\Lambda r_i^\Lambda = \sqrt{\Lambda}$ otherwise. In both cases $\Lambda r_i^\Lambda \le (r_i^\Lambda)^{\alpha_i}$. Hence,we can re-define $\epsilon(\delta)$ such that $\epsilon(\delta) \to 0$ as $\delta \to \infty$ and

$$\begin{aligned}\Lambda\bar{\Pi}_i^{\eta^\Lambda,P}(p) - \gamma_i\beta_i(\sqrt{R_i(p^*,0)}\eta_i^\Lambda) - \epsilon(\delta)(r_i^\Lambda)^{\alpha_i} \\ \le \bar{\Pi}_i^\Lambda(p,T) \le \\ \Lambda\bar{\Pi}_i^{\eta^\Lambda,P}(p) - \gamma_i\beta_i(\sqrt{R_i(p^*,0)}\eta_i^\Lambda) + \epsilon(\delta)(r_i^\Lambda)^{\alpha_i},\end{aligned} \tag{EC34}$$

for all $(p,T) \in \mathcal{L}^\Lambda(n^\Lambda)$. Also,

$$\bar{\Pi}_i^{\eta^\Lambda,P}(p) - \epsilon(\delta)(r_i^\Lambda)^{\alpha_i} \le \lambda_i(p,T)\left(p_i - c_i - \frac{\gamma_i}{\mu_i}\right) \le \bar{\Pi}_i^{\eta^\Lambda,P}(p) + \epsilon(\delta)(r_i^\Lambda)^{\alpha_i}. \tag{EC35}$$

We now apply a Tatônement scheme to both the market game and the fluid game on η^Λ and prove a coupling result. To that end, let $(p^{n^\Lambda+k,\Lambda}, T^{n^\Lambda+k,\Lambda})$ be the $n^\Lambda + k^{th}$ point in the trajectory of the market game under the Tatônement scheme for $k = 0, 1, 2, \ldots$. Let $\bar{p}^{\eta^\Lambda,k}$ be the k^{th} point in the trajectory of the fluid game on η^Λ when initialized at the point $p^{n^\Lambda,\Lambda}$ and under the corresponding Tatônement scheme. We then have the following result.

Lemma EC.4 *Given $\delta > 0$, there exists $\Lambda(\delta)$ and $\epsilon(\delta)$, such that, for all $i \in \mathcal{I}$,*

$$\left|\Pi_i^\Lambda(p^{n^\Lambda+k,\Lambda}, T^{n^\Lambda+k,\Lambda}) - \left(\Lambda\bar{\Pi}_i^{\eta^\Lambda,P}(\bar{p}^k) - \gamma_i\beta_i(\sqrt{R_i(p^*,0)}\eta_i^\Lambda)\right)\right| \le \Lambda\epsilon(\delta)(r_i^\Lambda)^{\alpha_i}, \tag{EC36}$$

and

$$|\bar{p}_i^{\eta^\Lambda,k} - p_i^{n^\Lambda+k,\Lambda}| \le M\sqrt{\epsilon(\delta)(r_i^\Lambda)^{\alpha_i}}, \tag{EC37}$$

for all $\Lambda \ge \Lambda(\delta)$, where M is the constant from Lemma 5.7. Moreover, $\epsilon(\delta) \to 0$ as $\delta \to 0$.

Using (EC34) and (EC35), the proof of Lemma EC.4 is very similar to that of Lemma EC.3 and we omit it. We now continue with the proof of Theorem C.1.

To this end, by the assumed global stability of the fluid game on η^Λ, we can choose $k^\Lambda(\delta)$ such that

$$|\bar{p}_i^{\eta^\Lambda,k} - p_i^*(\eta^\Lambda)| \le M\sqrt{\epsilon(\delta)(r_i^\Lambda)^{\alpha_i}},\ i \in \mathcal{I},$$

for all $k \geq k^\Lambda(\delta)$. Hence, for all $k \geq k^\Lambda(\delta)$,

$$|p_i^{k^\Lambda(\delta),\Lambda} - p_i^*(\eta^\Lambda)| \leq 2M\sqrt{\epsilon(\delta)(r_i^\Lambda)^{\alpha_i}},\ i \in \mathcal{I}.$$

Since the same holds for each $\delta > 0$, we can find sequences ϱ^Λ and $\tilde{n}^\Lambda \geq n^\Lambda$ such that $\varrho^\Lambda \to 0$ as $\Lambda \to \infty$ and so that$|p_i^{k,\Lambda} - p_i^*(\eta^\Lambda)| \leq M\sqrt{\epsilon(\delta)(r_i^\Lambda)^{\alpha_i}}$ for all $i \in \mathcal{I}$ and for all $k \geq \tilde{n}^\Lambda$. In particular,

$$\sup_{(p,T)\in\mathcal{L}^\Lambda(\tilde{n}^\Lambda)} |p_i - p_i^*(\eta^\Lambda)| \leq 2M\sqrt{\epsilon(\delta)(r_i^\Lambda)^{\alpha_i}},$$

for all Λ large enough and all $i \in \mathcal{I}$. Since $\delta > 0$ was arbitrary this concludes the proof of the theorem. ■

Proof of Lemma EC.3: The proof is by induction on k. For $k = 1$, equation (EC30) follows immediately from (EC29) and we turn to (EC31). As in (EC29) we have that for any $p_i \in \mathcal{P}_i$ and any T with $\|T\| \leq \delta$.

$$\left|\bar{\Pi}_i^P(p_{-i}^0, p_i) - \frac{\Pi_i^\Lambda(p_{-i}^0, p_i, T)}{\Lambda}\right| \leq \epsilon(\delta).$$

Assume, to reach a contradiction, that $|p_i^{1,\Lambda} - \bar{p}_i^1| > M\sqrt{\epsilon(\delta)}$. In particular, we have that

$$\left|\bar{\Pi}_i^P(p_{-i}^0, p_i^{1,\Lambda}) - \frac{\Pi_i^\Lambda(p_{-i}^0, p_i^{1,\Lambda}, T^{1,\Lambda})}{\Lambda}\right| \leq \epsilon(\delta) \text{ and } \left|\bar{\Pi}_i^P(p_{-i}^0, \bar{p}_i^1) - \frac{\Pi_i^\Lambda(p_{-i}^0, \bar{p}_i^1, T^{1,\Lambda})}{\Lambda}\right| \leq \epsilon(\delta).$$

Since $p_i^{1,\Lambda}$ and $\bar{p}_i^1$ are consequence of best responses, we have that

$$\begin{aligned}\bar{\Pi}_i^P(p_{-i}^0, p_i^{1,\Lambda}) + \epsilon(\delta) \geq \frac{\Pi_i^\Lambda(p_{-i}^0, p_i^{1,\Lambda}, T^{1,\Lambda})}{\Lambda} &\geq \frac{\Pi_i^\Lambda(p_{-i}^0, \bar{p}_i^1, T^{1,\Lambda})}{\Lambda} \\ &\geq \bar{\Pi}_i^P(p_{-i}^0, \bar{p}_i^1) - \epsilon(\delta) \geq \bar{\Pi}_i^P(p_{-i}^0, p_i^{1,\Lambda}) - \epsilon(\delta).\end{aligned}$$

Consequently,

$$\left|\bar{\Pi}_i^P(p_{-i}^0, \bar{p}_i^1) - \bar{\Pi}_i^P(p_{-i}^0, p_i^{1,\Lambda})\right| \leq \epsilon(\delta).$$

Since $\bar{p}_i^1$ is the best response to p_{-i}^0 in the fluid game we have by Lemma 5.7 that $|\bar{p}_i^1 - p_i^{1,\Lambda}| \leq M\sqrt{\epsilon(\delta)}$ and we have established both (EC30) and (EC31) for $k = 1$.

Assume now that (EC30) and (EC31) hold for $k = l - 1$ (with $l \geq 2$) and we will show it holds also for $k = l$. Assume that player i is the one that moves in the l^{th} round of the Tatônement scheme. As before, we have that

$$\left|\bar{\Pi}_i^P(p_{-i}^{l-1,\Lambda}, p_i^{l,\Lambda}) - \frac{\Pi_i^\Lambda(p_{-i}^{l-1,\Lambda}, p_i^{l,\Lambda}, T^{l,\Lambda})}{\Lambda}\right| \leq \epsilon(\delta),\ \left|\bar{\Pi}_i^P(p_{-i}^{l-1,\Lambda}, \bar{p}_i^1) - \frac{\Pi_i^\Lambda(p_{-i}^{l-1,\Lambda}, \bar{p}_i^1, T^{l,\Lambda})}{\Lambda}\right| \leq \epsilon(\delta),$$

$$\left|\bar{\Pi}_i^P(\bar{p}_{-i}^{l-1}, p_i^{l,\Lambda}) - \frac{\Pi_i^\Lambda(\bar{p}_{-i}^{l-1}, p_i^{l,\Lambda}, T^{l,\Lambda})}{\Lambda}\right| \leq \epsilon(\delta), \quad \left|\bar{\Pi}_i^P(\bar{p}_{-i}^{l-1}, \bar{p}_i^1) - \frac{\Pi_i^\Lambda(\bar{p}_{-i}^{l-1}, \bar{p}_i^1, T^{l,\Lambda})}{\Lambda}\right| \leq \epsilon(\delta).$$

Finally, since by the induction assumption $\max_{j\neq i} |p_j^{l-1,\Lambda} - \bar{p}_j^{l-1}\| \leq M\sqrt{\epsilon(\delta)}$ we also have that for all $p_i \in \mathcal{P}_i$,

$$\left|\bar{\Pi}_i^P(p_{-i}^{l-1,\Lambda}, p_i) - \bar{\Pi}_i^P(\bar{p}_{-i}^{l-1}, p_i)\right| \leq \epsilon(\delta).$$

From here the proof continuous as for $k = 1$, to show that

$$\left|\bar{\Pi}_i^P(\bar{p}_{-i}^{l-1}, \bar{p}_i^1) - \bar{\Pi}_i^P(\bar{p}_{-i}^{l-1}, p_i^{l,\Lambda})\right| \leq \epsilon(\delta),$$

so that by Lemma 5.7 we must have that $|p_i^{l,\Lambda} - \bar{p}_i^l| \leq M\sqrt{\epsilon(\delta)}$. Since $\max_{j\neq i} |p_j^{l,\Lambda} - \bar{p}_j^l| \leq M\sqrt{\epsilon(\delta)}$ by the induction assumption, this concludes the proof. ■