OPERATIONS RESEARCH
Vol. 58, No. 5, September–October 2010, pp. 1303–1317
ISSN 0030-364X | EISSN 1526-5463 | 10 | 5805 | 1303

DOI 10.1287/opre.1100.0827
© 2010 INFORMS

Investment and Market Structure in Industries with Congestion

Ramesh Johari
Department of Management Science and Engineering, Stanford University, Stanford, California 94305,
ramesh.johari@stanford.edu

Gabriel Y. Weintraub
Columbia Business School, New York, New York 10027, gweintraub@columbia.edu

Benjamin Van Roy
Department of Management Science and Engineering, Department of Electrical Engineering, Stanford University,
Stanford, California 94305, bvr@stanford.edu

We analyze investment incentives and market structure under oligopoly competition in industries with congestion effects. Our results are particularly focused on models inspired by modern technology-based services such as telecommunications and computing services. We consider situations where firms compete by simultaneously choosing prices and investments; increasing investment reduces the congestion disutility experienced by consumers. We define a notion of returns to investment, according to which congestion models inspired by delay exhibit increasing returns, whereas loss models exhibit nonincreasing returns. For a broad range of models with nonincreasing returns to investment, we characterize and establish uniqueness of pure-strategy Nash equilibrium. We also provide conditions for existence of pure-strategy Nash equilibrium. We extend our analysis to a model in which firms must additionally decide whether to enter the industry. Our theoretical results contribute to the basic understanding of competition in service industries and yield insight into business and policy considerations.

Subject classifications: competition; game theory; services; congestion.
Area of review: Manufacturing, Service, and Operations Management.
History: Received April 2009; revision received August 2009; accepted December 2009. Published online in *Articles in Advance* September 13, 2010.

1. Introduction

We consider oligopoly competition in service industries with *congestion effects*: the benefits consumers experience are offset by a negative externality that is increasing in the total volume of consumers served. Our base model consists of a finite collection of competing service providers facing a downward-sloping demand function. We consider a model where a consumer's disutility is a function of the *full price*: the sum of the price of the service and a congestion cost that increases with the total number of consumers subscribing to the same firm. Firms set prices, and also invest in their service; investment lowers the congestion cost experienced by their consumers.

Our model is motivated by modern technology-based services such as modern telecommunication and computing services; broadly, these services satisfy three key assumptions that drive our analysis. First, we assume that pricing and investment decisions are made on similar timescales; this will be the case in industries where investments are easily reversible. To capture this, we study a game where service providers choose prices and investment levels simultaneously, and consumers subsequently choose service providers. Second, we assume that consumers distribute among the firms so that the full prices of active firms are equalized; this will be the case if switching costs are relatively low for consumers. Third, we consider congestion models that include those where loss or blocking probability (as opposed to queueing delay) is the primary measure of disutility for consumers; more generally, the industries we consider are those that exhibit *nonincreasing returns to investment*. As we discuss below, these key assumptions regarding the nature of decisions and the cost structure fit well with important telecommunications and computing services, including wireless Internet service provision and cloud and cluster computing services.

In this paper, we make three main contributions regarding this class of oligopoly models. First, we characterize and study uniqueness and efficiency of Nash equilibrium in settings that exhibit nonincreasing returns to investment. Second, we study existence of equilibrium. Finally, we study a model where providers must first decide whether to enter the market. As we now discuss, our theoretical results contribute to the basic understanding of competition in service industries with congestion and provide insight into business and policy considerations.

We begin by defining a natural notion of *returns to investment.* We assume that the sum of congestion costs experienced by a firm's customers (called the *total congestion cost*) is jointly convex in the number of customers and the firm's investment expenditure. As a consequence of this fact, the industry exhibits nonincreasing returns to investment. The class of congestion cost models we consider accommodates *loss* sensitivity (e.g., where cost corresponds to the probability a job is dropped in a finite buffer queueing system) as a special case, but not *delay* sensitivity (e.g., where cost corresponds to delay in an infinite buffer queueing system); delay models exhibit increasing returns to investment. Our results establish that the nature of congestion sensitivity (loss versus delay) has a first-order impact on market structure.

Next, we study uniqueness and efficiency of pure-strategy Nash equilibria of the simultaneous pricing and investment game. First, we consider an industry that exhibits constant returns to investment. We show that in this case the *total cost* of each firm (i.e., the sum of investment and congestion costs when a firm invests efficiently) is linear in the total demand served in equilibrium, greatly simplifying the analysis. We show that every Nash equilibrium has a threshold form: a firm is active if and only if the slope of its total cost is below a threshold. Moreover, we show that if such an equilibrium exists, it is unique. We then consider an industry that exhibits nonincreasing returns to investment, but in which firms are homogeneous (i.e., they all share the same congestion cost function). For this model, we prove that if a pure-strategy Nash equilibrium exists, it is unique and symmetric (i.e., all firms are active).

Our uniqueness results provide a sharp characterization of equilibrium behavior. In particular, our results allow us to study the effects of both demand elasticity and heterogeneity among firms on efficiency of the resulting equilibrium. As long as all firms are homogeneous, we show that the unique Nash equilibrium is efficient conditional on the total number of consumers served; however, because firms have market power, the number of consumers served is below the socially efficient level. As demand becomes perfectly inelastic, the unique Nash equilibrium becomes efficient. We also observe via numerical example that as firms become increasingly heterogeneous, inefficiency can increase significantly as well. In this situation it is possible for an efficient firm to price less efficient firms out of the market and yet realize an operating point that exhibits significant inefficiency.

Pure-strategy Nash equilibrium may not exist in general, so we then provide sufficient conditions for their existence. In particular, we observe that if the congestion cost is "too steep" with respect to the number of firms in the industry, a pure-strategy Nash equilibrium may fail to exist in a model with perfectly inelastic demand. Motivated by this negative result, we provide several distinct precise conditions that guarantee existence of pure-strategy Nash equilibrium. We begin by showing that if the demand and the congestion cost functions are concave, a pure-strategy Nash equilibrium exists. We also provide sufficient conditions for existence of Nash equilibrium in settings with constant returns to investment and an elastic demand curve, and with nonincreasing returns to investment and an inelastic demand curve. In both these cases, we require that the congestion cost is not too steep relative to the number of firms in the industry.

The preceding results pertain to competition among a given number of incumbent firms. However, the number of participating firms has a significant impact on market performance. With this motivation, we extend our analysis to include an *entry stage*, and we study the efficiency properties of entry decisions made by homogeneous profit-maximizing firms. In that analysis, we assume that entrants pay a positive fixed sunk cost to compete, and the industry exhibits constant returns to investment. We establish that the equilibrium number of entrants exceeds the socially efficient level; however, entry becomes efficient asymptotically as the sunk entry cost becomes small. We also study entry decisions in an industry that exhibits nonincreasing returns to investment, but that faces a perfectly inelastic demand.

Our uniqueness, existence, and entry results are the first for the class of models we study. Extensive attention has been devoted to analyzing oligopoly models with congestion in the recent literature in operations, economics, communication networks, and transportation; see, for example, Xiao et al. (2007), Allon and Federgruen (2008), Acemoglu et al. (2009), Acemoglu and Ozdaglar (2007), Cachon and Harker (2002), Scotchmer (1985), and the detailed discussion in §2. As several of these authors have noted, important basic features of such models have remained poorly understood, particularly concerning existence and uniqueness of equilibrium. The presence of congestion effects distinguishes these models from standard price-setting or quantity-setting oligopoly games (Vives 2001). Moreover, equilibrium analysis is especially difficult in the case of games where firms choose both prices and investment levels, because such games are generally neither concave nor supermodular—thus, standard game-theoretic arguments do not apply. Therefore, our work represents a significant contribution to the basic understanding of competition in congested industries.

We conclude by discussing the implications of our results for policy analysis and business strategy. Our model and analysis are directly relevant for modern technology-based services such as modern telecommunication and computing services. As we now establish, such industries satisfy our key assumptions regarding the timing of decisions, customer behavior, and returns to investment. In particular, in these industries (1) pricing and investment in capacity can be carried out on a similar timescale, (2) consumers have relatively low switching costs, and (3) constant returns to investment are exhibited. As a result, our insights provide a benchmark with which a range of such service industries can be studied.

First, consider a wireless hotspot (Wi-Fi) provider who offers Internet access to consumers. We assume that the provider can invest in additional wireless access points (AP) to expand the capacity of the network, but that at each access point the number of channels available for transmission is constrained by the Wi-Fi protocol and available spectrum. Consumers are sensitive to channel access congestion, as measured by the experienced loss (or blocking) probability when they try to use the service. Because APs are inexpensive and easily installed, capacity planning can be carried out on the same timescale as service pricing. (Indeed, infrastructure solutions such as those offered by Meraki, www.meraki.com, have vastly simplified the process of expanding hotspot capacity.) For consumers, switching costs are typically quite low between different hotspots—often as simple as choosing an alternative hotspot via a software interface. Furthermore, a simple channel access congestion model analogous to that studied by Campo-Rembado and Sundararajan (2004) suggests that this industry exhibits constant returns to investment.

Second, consider the rapidly growing cloud computing platforms, such as the Force.com service offered by Salesforce.com (www.salesforce.com/force), and the cloud computing services offered by Amazon's Elastic Compute Cloud (EC2) service (aws.amazon.com/ec2), GoGrid (www.gogrid.com), and Flexiscale (www.flexiscale.com). These services aggregate large amounts of computing resources into clusters and employ sophisticated resource allocation mechanisms to sell "virtual" computers created from these resources. Such services allow nascent software developers to rapidly scale up their platforms without needing the large capital investment of building their own computing cluster. We consider a model where the provider has already made the large capital investment to establish several geographically dispersed computing clusters; in this case, investment is primarily in the computing hardware and network connectivity available within each cluster. These are easily upgradeable commodity elements that can be altered on the same timescale as prices. From a customer's standpoint, switching costs are relatively low, precisely because these services are virtual computing platforms: for example, Flexiscale even advertises "true pay-as-you-go utility pricing with no lock-in." Furthermore, in cloud services, applications are subject to blocking if resources are not available—again a congestion model that satisfies constant returns to investment.

Our results suggest that for these industries if all firms have access to the same technology, and, hence, they are homogeneous, competition yields outcomes that are socially desirable; the unique equilibrium is symmetric and no dominant firm emerges. Moreover, firms invest efficiently conditional on the number of consumers they serve. These appealing properties are not obtained in situations with increasing returns to investment or where investments are chosen before prices. In the former, a natural monopoly arises, and in the latter, firms underinvest to soften price competition. On the other hand, if some firms have technological advantages over others, then even under our assumptions on timing and returns to investment, the efficiency loss compared to a model with homogeneous firms is generally larger, and firms with cost advantages can exploit their market power and price less efficient firms out of the market.

The remainder of the paper is organized as follows. In §2 we review literature related to our work. In §3 we introduce our model of service provision. In §4 we introduce and study the notion of returns to investment. In §5 we introduce a game-theoretic model to analyze competition between profit-maximizing firms, and characterize its Nash equilibrium. In §6 we study uniqueness of Nash equilibrium. In §7 we study existence of equilibria. In §8 we study entry decisions made by firms. Finally, in §9 we conclude and provide some thoughts for future research. We note that all appendices, including proofs, are provided in the e-companion to this paper, which is available as part of the online version that can be found at http://or.journal.informs.org/.

2. Related Literature

In this section we briefly discuss several threads of the literature related to our model and analysis. We compare our work to recent results in the operations management literature and to welfare analysis in congestion games. We also discuss models of Edgeworth-Bertrand games and club goods.

Several operations management papers study competition in service industries (see Allon and Federgruen 2008, Cachon and Harker 2002, So 2000, Allon and Federgruen 2007, for a survey). In these studies congestion models are often based on the steady-state expected waiting time of a typical customer in a queue. In these models, resource pooling is typically efficient, so in the context of our model, there are increasing returns to investment. Furthermore, the game-theoretic model is substantially different: firms commit ex ante to a guaranteed level of service and invest ex post to meet that guarantee. DiPalantino et al. (2009) compare the model in this paper with a service-level guarantee model in terms of market outcomes and show that equilibria can be drastically different.

Our paper is related to the growing recent literature on welfare analysis in congestion games in transportation and communication networks; see, e.g., Roughgarden (2005) for an overview. In particular, Acemoglu and Ozdaglar (2007), Ozdaglar (2008), Hayrapetyan et al. (2007), and Engel et al. (2004) study competition among profit-maximizing oligopolists that set prices, whereas consumers' disutility is measured through the sum of price and congestion cost (as in our paper). Our paper extends their analysis by including investment and entry decisions.

Closely related to our work is Xiao et al. (2007), which independently simultaneously studied a pricing and investment game similar to the one studied in this paper. Their entire analysis is restricted to industries that exhibit constant returns to investment only, whereas part of our analysis includes a wide range of industries that exhibit nonincreasing returns. A main focus of their paper is to bound efficiency loss of a pure-strategy Nash equilibrium. Specific bounds are obtained for *symmetric* equilibria when firms are assumed to be homogeneous. They found that the efficiency loss of symmetric Nash equilibrium (NE) compared to the social optimum is no more than 15% for exponential or linear demand functions. This result complements our insights; efficiency loss of symmetric NE is not significant in these cases. However, Xiao et al. (2007) do not study uniqueness and existence of Nash equilibria. Hence, our results strengthen theirs, because our uniqueness result implies that their bounds for symmetric equilibria are valid even if one considers asymmetric equilibria. Furthermore, our existence result ensures that these bounds are not vacuous. We conclude by noting that Xiao et al. (2007) assume that in the socially optimal solution all firms are active. In a setting with constant returns to investment, this is only possible if all firms share the *same* total cost function. By contrast, we analyze uniqueness and existence of equilibrium among firms with heterogeneous total cost functions; in our approach we explicitly consider participation, which is a fundamental determinant of both socially efficient and equilibrium outcomes.

The presence of congestion effects distinguishes our model from standard pricing- or quantity-setting oligopoly games (Vives 2001). Our model is more closely related, however, to Edgeworth-Bertrand games, where firms face strict capacity constraints and compete by setting prices (Edgeworth 1925, Tirole 1988). In an Edgeworth-Bertrand game, no congestion is experienced until capacity is reached, and congestion is “infinite” thereafter. In contrast, in our model, congestion is monotonically increasing in the number of consumers. In Edgeworth-Bertrand games, when firms compete by setting quantities and prices simultaneously, pure-strategy Nash equilibria generally do not exist (Levitan and Shubik 1978) unless demand is stochastic and the game is *large* (Deneckere and Peck 1995). This is in marked contrast to our results. In related work, Acemoglu et al. (2009) show that existence of equilibrium can also be restored if capacity decisions are made prior to pricing decisions.

Our paper is also closely related to the literature on “club goods” from public economics, which analyzes shared public goods with congestion, such as swimming pools (see Scotchmer 2002 for a recent survey). In particular, Scotchmer (1985) studies Nash equilibrium among profit-maximizing clubs that first choose whether to enter, and then choose facility size and price simultaneously at the second stage given a perfectly inelastic demand. Our analysis is more general than Scotchmer's model because her entire analysis is restricted to a perfectly inelastic demand and homogeneous firms. In addition, we prove uniqueness and existence of a pure-strategy Nash equilibrium; Scotchmer (1985) does not study uniqueness and proves existence for large economies only. De Vany and Saving (1983) analyze a similar model but consider competitive equilibria.

Like our uniqueness and existence results for the pricing and investment game, our entry results are also the first for the class of models we study. Xiao et al. (2007) do not study entry; Scotchmer (1985) studies entry in a similar model, but that analysis does not apply to our model. As we discuss later in §8, Scotchmer's entry results are vacuous in our setting, because she assumes the sunk entry cost equals zero. Our entry results extend classic results in standard oligopoly games (Mankiw and Whinston 1986) to congestion games.

3. Model

In this section we introduce our model of service provision, focusing on competition *after* firms have already entered the market. We assume that $N \geqslant 2$ incumbent firms are present after entry decisions have been made; we consider a game with entry decisions in §8. Firms compete for consumers by choosing prices and investment levels. Investment made by a firm improves the service experience for all consumers that are served by that firm. We assume there are no externalities among firms; therefore, consumers served by other firms are unaffected by this investment.

For firm j, we let p_j, I_j, and x_j denote, respectively, the price per consumer charged, the investment level chosen, and the number of consumers served. Each investment level I_j is measured in currency units, and the resulting physical capacity can be a nonlinear function of this investment expenditure. The postentry profit of firm j is given by:[1]

$$\pi(p_j, I_j, x_j) = p_j x_j - I_j. \tag{1}$$

Thus, profits of firms are determined by the price, investment expenditure, and number of consumers served; of these, price and investment expenditure are decision variables for the firms.

The demand model formalizes a congestion externality among a firm's consumers. We assume that when a firm j invests I_j and serves x_j consumers, each consumer of that firm experiences a *congestion cost* $l_j(x_j, I_j)$. The congestion cost function $l_j(x_j, I_j)$ represents the disutility perceived by consumers due to congestion.

Assumption 1. *For each j, the congestion cost function $l_j(x_j, I_j)$ is finite for all $x_j \geqslant 0$ and $I_j > 0$, and is twice differentiable in this region. Furthermore, for all $x_j > 0$ and $I_j > 0$, $\partial l_j(x_j, I_j)/\partial x_j > 0$, $\partial l_j(x_j, I_j)/\partial I_j < 0$, and $l_j(0, I_j) = 0$. In addition, $l_j(0, 0) = \infty$, and $\lim_{I_j \downarrow 0} l_j(x_j, I_j) = l_j(x_j, 0) = \infty$ for all $x_j > 0$.*

The assumption implies that congestion increases with the mass of subscribers and decreases with investment expenditures. The assumption also incorporates natural boundary conditions: there is no congestion if there are no subscribers, and infinite congestion cost if a service provider retains subscribers but does not invest in any capacity. *Assumption* 1 *is maintained throughout the paper unless otherwise explicitly noted.*

We assume that congestion cost is measured in currency equivalent terms. Hence, a customer's utility depends on the sum of the price he is charged for service and the congestion cost he experiences. We call this sum the *full price*.

Definition 1. The *full price* experienced by a customer of firm j is equal to $p_j + l_j(x_j, I_j)$.

Consumers generate a downward-sloping demand function. We let $D(\Delta)$ denote the demand function, and let $P(q)$ be the inverse demand function; i.e., $P(D(\Delta)) = \Delta$ for all $\Delta > 0$. We interpret $P(q)$ as the marginal utility obtained by an additional infinitesimal consumer when the total number of consumers being served is q. We make the following standard assumption *that is maintained throughout the paper unless otherwise explicitly noted.*

Assumption 2. *For all* $q \geqslant 0$, $P(q)$ *is nonnegative and continuously differentiable with* $P'(q) < 0$ *whenever positive. Furthermore,* $\lim_{q\to\infty} P(q) = 0$.[2]

To model consumer behavior, we consider a static equilibrium in which firms that attract customers offer the same full price. This is a natural condition: if one firm offers a higher full price than another, then, absent switching costs, its customers would switch providers. Such an equilibrium is commonly known as a *Wardrop equilibrium*, particularly in the transportation literature; we adopt the same terminology here, with the abbreviation WE (Wardrop 1952). WE is commonly used in this class of congestion models (e.g., Roughgarden 2005, Acemoglu and Ozdaglar 2007, Engel et al. 2004). Formally, we have the following definition. We use boldface type to denote vectors.

Definition 2. For given price and investment vectors $\mathbf{p}$ and $\mathbf{I}$, a vector of demand quantities $\mathbf{x} \geqslant 0$ is a *Wardrop equilibrium* if

$$p_j + l_j(x_j, I_j) = P(Q), \quad \text{for all } j \text{ with } x_j > 0; \tag{2}$$

$$p_j + l_j(x_j, I_j) \geqslant P(Q), \quad \text{for all } j, \tag{3}$$

where $Q = \sum_{i=1}^N x_i$.

Under Assumptions 1 and 2, given price and investment vectors $\mathbf{p}$ and $\mathbf{I}$, if $I_j > 0$ for at least one firm j, then a WE exists and is unique; see, e.g., Beckmann et al. (1956). We denote the set of WE by $W(\mathbf{p}, \mathbf{I})$. When $I_j = 0$ for all j, we let $W(\mathbf{p}, \mathbf{I}) = \varnothing$.

We now introduce the problem a social planner would solve. The solution of this problem, which we call the *efficient solution*, provides a benchmark against which equilibrium outcomes will be compared. We consider the problem of maximizing *social surplus*, defined as the sum of consumer and producer surplus, *given* a fixed number $N \geqslant 2$ of incumbent firms.[3]

Definition 3. The pair of vectors $\mathbf{x}^S$ and $\mathbf{I}^S$ is a *social optimum*, or *efficient solution*, if it maximizes total social surplus; i.e., if it solves:

$$\text{maximize} \int_0^{\sum_{i=1}^N x_i} P(q)\, dq - \sum_{i=1}^N (x_i l_i(x_i, I_i) + I_i) \tag{4}$$

subject to $\mathbf{x}, \mathbf{I} \geqslant 0$.

We define the socially optimal mass of consumers served Q^S according to $Q^S = \sum_{j=1}^N x_j^S$.

For later reference, we define efficient investment and the total cost function.

Definition 4. Given total customer mass $x_j \geqslant 0$, a firm j's *efficient investment level* $I_j(x_j)$ is an investment level that minimizes the sum of total congestion cost and investment cost. That is, $I_j(x_j)$ is a minimizer of the following optimization problem:[4]

$$v_j(x_j) \equiv \min_{I_j \geqslant 0} [x_j l_j(x_j, I_j) + I_j]. \tag{5}$$

The function v_j is called the *total cost function* for firm j.

Observe that because investment must be efficient at the socially optimal solution, the social planner's problem is equivalent to maximizing $\int_0^Q P(q)\, dq - \sum_i v_i(x_i)$ over $\mathbf{x} \geqslant 0$, where $Q = \sum_i x_i$. We will use this characterization to compare Nash equilibrium outcomes with socially optimal outcomes.

Finally, we make the following standard assumption *that is maintained throughout the paper unless otherwise explicitly noted*; without this assumption, no firm would serve any customers in either the socially efficient solution or in equilibrium.[5]

Assumption 3. $P(0) > \min_i \lim_{x_i \to 0} v_i'(x_i)$.

4. Returns to Investment

Cost structure is a key determinant of market outcomes in our model. In this section, we define a notion of *returns to investment* that yields a unifying framework for classifying cost structures arising from different congestion models. We start with the following definition; we use $l(x, I)$ to generically refer to the congestion cost function of a given firm.

Definition 5. The *total congestion cost* experienced by a mass x of customers served by a firm with congestion cost function $l(x, I)$ that invests I is $K(x, I) = xl(x, I)$.

Recall that $l(x, I)$ represents the congestion cost experienced per unit mass of customers. Hence, $K(x, I)$ represents the sum of congestion costs experienced by all subscribers to a firm's service.

Returns to investment are defined via the total congestion cost function $K(x, I)$ as follows.

Definition 6. A firm with congestion cost function $l(x, I)$ exhibits *nonincreasing* (*respectively, nondecreasing*) *returns to investment* if:

$$K(\alpha x, \alpha I) \geqslant (\text{resp., } \leqslant)\ \alpha K(x, I), \qquad \text{for all } \alpha > 1, \text{ and } x, I > 0.$$

The firm exhibits *decreasing* (*increasing*) *returns to investment* if the corresponding inequalities are strict. The firm exhibits *constant returns to investment* if returns to investment are both nonincreasing and nondecreasing.

To develop intuition, consider a setting where all firms share the same congestion cost function. If firms exhibit increasing returns to investment, then given a fixed investment expenditure, the total congestion cost associated with a single firm serving the entire market is smaller than the cost associated with several firms splitting both the demand and the investment expenditure equally. If firms exhibit decreasing returns to investment, then the converse is true.[6]

In this paper, *we focus primarily on models that exhibit nonincreasing returns to investment*. Formally, we consider congestion cost functions that satisfy the following convexity assumption.

Assumption 4. *For all j, the total congestion cost $K_j(x_j, I_j) = x_j\, l_j(x_j, I_j)$ is jointly convex in (x_j, I_j), and strictly convex in I_j, for each $x_j > 0$.*[7]

It is straightforward to verify that if the total congestion cost K is convex, then a firm with congestion cost l exhibits *nonincreasing* returns to investment. If Assumption 4 holds, then at any social optimum a positive mass of consumers is served and investment is efficient. Moreover, the total cost functions v_j can be shown to be convex; see Lemma EC.3 in Appendix EC.2. Finally, the optimal solution of problem (5) is also unique in this case; that is, for all j and $x_j \geqslant 0$, the efficient investment level $I_j(x_j)$ is unique.

For several of our results we will assume firms exhibit constant returns to investment, as follows.

Assumption 5. *Assumption* 4 *holds. Moreover, for all j, firm j exhibits constant returns to investment*; *that is, there exists a function h_j such that $l_j(x, I) = h_j(x/I)$.*

If Assumption 5 holds, then v_j is in fact *linear*. For simplicity, we omit the dependence of l, I, and v on the subscript j.

Lemma 1. *Suppose Assumption* 5 *holds for congestion cost function l*; *i.e., there exists a function h such that $l(x, I) = h(x/I)$. Then there exists a unique solution ϕ to the equation $\phi^2 h'(\phi) = 1$. Furthermore, $I(x) = x/\phi$, and thus $v(x) = \xi x$, where $\xi = h(\phi) + 1/\phi > 0$.*

Several key examples satisfy Assumption 5. These include service systems that can be modeled as loss systems (i.e., consumers' disutility is a function of the blocking probability) and where firms invest to increase the service rate. As one example, Hall and Porteus (2000) use loss system models to analyze competition in capacitated systems. Xiao et al. (2007) use a similar model to analyze competition among private toll roads. Loss systems are also a plausible model for wireless service provision, where we expect that consumers are most sensitive to the fraction of times they are unable to connect to a base station after paying a subscription fee to a given provider.[8] Indeed, constant returns to investment are exhibited, for example, when the marginal productivity of investment expenditure in building capacity is constant and $l(q, I)$ represents Erlang's formula for a loss system with mean arrival rate q, service rate I, and a fixed number of servers. Constant returns to investment are also exhibited for alternative loss models like the loss probability of an $M/M/1/s$ system or the exceedance probability of an $M/M/1$ queue (Kleinrock 1975). If the marginal productivity of investment expenditure in building capacity is decreasing, then these models exhibit decreasing returns to investment. In Appendix EC.1 we provide details on these and other related examples and prove that they satisfy our assumptions.

We conclude this section by briefly considering the scenario where an industry exhibits increasing returns to investment; note that in this case K is not convex, i.e., Assumption 4 is not satisfied. In this setting we typically expect that the efficient solution calls for a single firm serving the entire market and that a natural monopoly will arise. As noted in Appendix EC.1, an important class of congestion models with this property is derived from steady-state expected waiting times in queueing models.

A key resulting insight is that the nature of the congestion cost experienced by consumers is a primary determinant of the returns to investment in the industry. In turn, the returns to investment have a fundamental impact on efficiency of market outcomes. As illustrated by the examples in Appendix EC.1, industries where consumers are *loss sensitive* exhibit *nonincreasing returns* to investment, whereas industries where consumers are *delay sensitive* exhibit increasing returns. Hence, the distinction between delay and loss is critical for market outcomes. As noted in the introduction, our emphasis is on technology-based services, including wireless Internet service provision and cloud computing services; these services fit well with the assumption of constant or nonincreasing returns to investment.

5. The Game and Nash Equilibrium

In this section we introduce a game-theoretic model to analyze competition between profit-maximizing firms. We analyze the postentry game with a fixed finite number of incumbent firms N. We consider a game where prices and investment levels are chosen simultaneously; it is as if the two decisions are made on the same timescale, and investment decisions are as reversible as pricing decisions. We first define and characterize pure-strategy Nash equilibrium in prices and investment levels. Then, using this characterization, in §§6 and 7 we prove several of the main results of the paper: these establish uniqueness, existence, and efficiency properties of Nash equilibrium.

We study *pure-strategy Nash equilibrium* (NE) of the simultaneous pricing and investment game, defined as follows.

Definition 7. A triple consisting of prices $\mathbf{p}^{\mathrm{NE}}$, investment levels $\mathbf{I}^{\mathrm{NE}}$, and demand quantities $\mathbf{x}^{\mathrm{NE}}$, is a *pure-strategy Nash equilibrium* of the simultaneous pricing and investment game (NE) if the following conditions hold:

1. The demand quantities are a WE given prices and investment levels: $\mathbf{x}^{\mathrm{NE}} \in W(\mathbf{p}^{\mathrm{NE}}, \mathbf{I}^{\mathrm{NE}})$.

2. Each firm maximizes profit given prices and investment levels of other firms; i.e., for all $j = 1, \ldots, N$, $p_j, I_j \geqslant 0$, and $\mathbf{x} \in W(p_j, \mathbf{p}_{-j}^{\mathrm{NE}}, I_j, \mathbf{I}_{-j}^{\mathrm{NE}})$,

$$\pi(p_j^{\mathrm{NE}}, I_j^{\mathrm{NE}}, x_j^{\mathrm{NE}}) \geqslant \pi(p_j, I_j, x_j). \tag{6}$$

Note that when a firm makes investment and pricing decisions, it anticipates that consumers will be allocated according to a WE. We use $\mathbf{p}_{-j}$ and $\mathbf{I}_{-j}$ to denote the vectors of prices and investment levels of the competitors of firm j.

In §5.1, we provide Nash equilibrium conditions for the general model. In §5.2, we specialize the conditions to a setting where firms are homogeneous (i.e., share the same congestion cost characteristics); in this case our interest will be in *symmetric* equilibrium.

5.1. Nash Equilibrium Conditions: Heterogeneous Firms

We first find necessary conditions for a Nash equilibrium in the general model, where firms may be heterogeneous. For our development we require the concept of an active firm.

Definition 8. A firm j is *active* at a NE $(\mathbf{p}^{\mathrm{NE}}, \mathbf{I}^{\mathrm{NE}}, \mathbf{x}^{\mathrm{NE}})$ if it invests a positive amount $I_j^{\mathrm{NE}} > 0$.

Note that in equilibrium, only active firms serve customers. We establish the following proposition. Xiao et al. (2007) proves a similar result for the special case of industries that exhibit constant returns to investment. The proof is provided in Appendix EC.3.

Proposition 1. *Suppose that the vectors of prices $\mathbf{p}^{NE}$, investment levels $\mathbf{I}^{NE}$, and demand levels $\mathbf{x}^{NE}$ form an NE for which only firms in the set A are active, where A is a nonempty subset of $\{1, 2, \ldots, N\}$. Then the NE must satisfy the following conditions*:

$$p_j^{NE} = x_j^{NE}\Bigg(\frac{\partial l_j(x_j^{NE}, I_j^{NE})}{\partial x_j} + \frac{1}{\sum_{i\in A: i\neq j}(1/(\partial l_i(x_i^{NE}, I_i^{NE})/\partial x_i)) - 1/(P'(Q^{NE}))}\Bigg), \quad j \in A; \tag{7}$$

$$0 = x_j^{NE}\frac{\partial l_j(x_j^{NE}, I_j^{NE})}{\partial I_j} + 1, \quad j \in A, \tag{8}$$

where $Q^{NE} = \sum_{j=1}^{N} x_j^{NE}$. Furthermore, $P(Q^{NE}) > 0$; hence, $P'(Q^{NE}) < 0$.

In the NE, firm $j \in A$ makes a profit equal to $\pi(p_j^{NE}, I_j^{NE}, x_j^{NE}) = P(Q^{NE})x_j^{NE} - v_j(x_j^{NE})$. Furthermore, if Assumption 4 also holds, then all firms invest efficiently: $I_j^{NE} = I_j(x_j^{NE})$ *for all firms $j \in A$.*

Efficient investment follows because under Assumption 4, (8) is the optimality condition for (5). Intuitively, firms invest at efficient levels because they can extract any additional consumer surplus generated by investment through an appropriate choice of price. This insight was also previously obtained by Scotchmer (1985) for club goods.

Note that if a social planner were to levy "taxes" to induce a social optimum $(\mathbf{x}^S, \mathbf{I}^S)$, she should charge a *Pigovian price* for the service of each firm j, given by $p_j = x_j^S \partial l_j(x_j^S, I_j^S)/\partial x_j$ (Pigou 1920). This corresponds to the congestion externality imposed by the marginal consumer at firm j to all other consumers served by firm j. The NE price p_j^{NE} is the Pigovian price plus a positive *markup*. The price reflects the fact that firm j charges an additional marginal customer the amount required to retain existing consumers. A new marginal customer imposes a congestion externality on existing customers. In addition, the marginal unit of demand is partially derived from the competitors; hence, their congestion levels are reduced. Firm j needs to compensate its customers for these two factors to retain them despite its higher congestion.

5.2. Nash Equilibrium Conditions: Homogeneous Firms

In several of our results, we consider a specialized setting where firms are *homogeneous*; i.e., they share the same congestion cost specification. This is formalized in the following assumption.

Assumption 6. *All firms have the same congestion cost function*: *for all x, I and for all i, j, there holds $l_i(x, I) = l_j(x, I)$.*

Whenever Assumption 6 holds, we suppress subscripts on the functions l, h, K, I, and v.

With homogeneous firms, we are particularly interested in symmetric NE, defined as follows.

Definition 9. An NE is *symmetric* if $p_i^{\mathrm{NE}} = p_j^{\mathrm{NE}}$ and $I_i^{\mathrm{NE}} = I_j^{\mathrm{NE}}$ for all i, j; because the WE is uniquely defined, this also implies $x_i^{\mathrm{NE}} = x_j^{\mathrm{NE}}$. An NE is *symmetric among active firms* if $p_i^{\mathrm{NE}} = p_j^{\mathrm{NE}}$ and $I_i^{\mathrm{NE}} = I_j^{\mathrm{NE}}$ for all firms i, j that are active.

Note that if Assumptions 4 and 6 hold, then there exists a *symmetric* social optimum. In this setting, symmetry of NE is a socially desirable property, because in that case both demand allocations and investment levels are efficient conditional on the total mass of consumers served.

When firms are homogeneous, the necessary condition for a symmetric NE becomes:

$$p_j^{\mathrm{NE}} = x_j^{\mathrm{NE}} \left(\frac{\partial l(x_j^{\mathrm{NE}}, I_j^{\mathrm{NE}})}{\partial x_j} + \frac{1}{(N-1)/(\partial l(x_j^{\mathrm{NE}}, I_j^{\mathrm{NE}})/\partial x_j) - 1/(P'(Q^{\mathrm{NE}}))} \right), \quad \text{for all } j. \tag{9}$$

Finally, we conclude by specializing further to a setting where demand is perfectly *inelastic*.

Assumption 7. *Demand is perfectly inelastic of size M.*

The assumption corresponds to a situation where a total customer mass of size M has an infinite valuation for the service. In Appendix EC.4, we formally develop the model of perfectly inelastic demand with homogeneous firms and the resulting Wardrop equilibrium conditions for demand allocation. In this case, both in the social optimum and NE, the entire mass of consumers M is served. It can be shown that the price at a symmetric NE is given by:

$$p_j^{\mathrm{NE}} = \frac{M}{N-1} \frac{\partial l}{\partial x}\left(\frac{M}{N}, I_j^{\mathrm{NE}}\right). \tag{10}$$

We observed above that when firms are homogeneous and Assumption 4 holds, demand allocations and investment levels are efficient conditional on the total mass of consumers served. When demand is inelastic, this property implies the additional insight that a symmetric NE is in fact *socially efficient*.

6. Uniqueness of Nash Equilibrium

In this section we prove several of the main results of the paper, concerning uniqueness and efficiency of NE in the oligopolistic simultaneous pricing and investment game. In §6.1 we consider a class of models that exhibits constant returns to investment; we show that if a NE exists, then it is unique. In §6.2 we assume that firms are homogeneous, and consider a class of models that exhibit nonincreasing returns to investment; we show that if an NE exists, then it is unique and symmetric. Moreover, if demand is perfectly inelastic, this NE is efficient. These results provide a sharp characterization of NE behavior. In §6.3 we use our results to discuss the implications of NE behavior in terms of social welfare.

6.1. Uniqueness: Heterogeneous Firms and Constant Returns to Investment

In this section, we show that if firms exhibit constant returns to investment (i.e., if Assumption 5 holds), then if a NE exists, it is unique. Recall that under Assumption 5, for all j, the total cost function v_j is linear; i.e., $v_j(x) = \xi_j x$ for some $\xi_j > 0$ (see Lemma 1). Linearity of the total cost function greatly simplifies the analysis. Without loss of generality, *for the remainder of the paper whenever Assumption 5 holds, we also assume that $\xi_1 \leqslant \cdots \leqslant \xi_N$.* That is, firm 1 has the lowest total cost function and firm N has the highest. Recall that in Appendix EC.1 we discuss important congestion cost functions that exhibit constant returns to investment.

The following result shows that an NE must be of a threshold form: all firms with cost coefficient below a threshold are active, and all others are not. All proofs for this section are provided in Appendix EC.5.

Proposition 2. *Suppose Assumption 5 holds. Suppose that the vectors of prices $\mathbf{p}^{NE}$, investment levels $\mathbf{I}^{NE}$, and demand levels $\mathbf{x}^{NE}$ form an NE with at least one active firm. Let $Q^{NE} = \sum_{j=1}^{N} x_j^{NE}$. Then, the profit of firm j is given by $p_j^{NE} x_j^{NE} - I_j^{NE} = (P(Q^{NE}) - \xi_j) x_j^{NE}$. Moreover, firm j is active if and only if $P(Q^{NE}) > \xi_j$. As a consequence, the NE is a threshold equilibrium: there exists $n^* \in \{1, \ldots, N\}$ such that all firms $i \leqslant n^*$ are active, and all firms $i > n^*$ are not active.*

Next, we show that for any fixed threshold n^*, there is essentially at most one NE with that threshold, as long as demand is log concave. (A positive function f is log concave if $\log f$ is concave.) Recall that $D(\Delta)$ is the demand function, i.e., $P(D(\Delta)) = \Delta$ for all $\Delta > 0$. Note that many commonly used demand functions are log concave, such as $D(\Delta) = \exp(-\Delta)$.

Proposition 3. *Suppose Assumption 5 holds and the demand function $D(\Delta)$ is log concave over the region where it is positive. Fix $n^* \geqslant 1$. Suppose that the vectors of prices $\mathbf{p}^{NE}$, investment levels $\mathbf{I}^{NE}$, and demand levels $\mathbf{x}^{NE}$ form an NE with n^* active firms. Then, $\mathbf{x}^{NE}$ and $\mathbf{I}^{NE}$ are uniquely determined; further, prices for active firms, p_j^{NE}, $j = 1, \ldots, n^*$, are uniquely determined as well.*

The preceding proposition establishes that for any fixed threshold, there is at most one NE with that threshold. We now show that the threshold for any NE is uniquely determined as well.

Proposition 4. *Suppose that Assumption 5 holds and the demand function $D(\Delta)$ is log concave over the region where it is positive. Suppose that the vectors of prices $\mathbf{p}^{NE}$, investment levels $\mathbf{I}^{NE}$, and demand levels $\mathbf{x}^{NE}$ form an NE with n^* active firms. Then, $n^* \geqslant 1$ and n^* is uniquely determined.*

The previous results lead directly to the main result of this section: if demand is log concave, then the NE is essentially uniquely determined.

Theorem 1. *Suppose Assumption 5 holds and the demand function $D(\Delta)$ is log concave over the region where it is positive. Suppose that the vectors of prices $\mathbf{p}^{NE}$, investment levels $\mathbf{I}^{NE}$, and demand levels $\mathbf{x}^{NE}$ form an NE; let n^* be the number of active firms, and let $Q^{NE} = \sum_{j=1}^{N} x_j^{NE}$.*

Then $n^* \geqslant 1$, *and* n^*, $\mathbf{x}^{NE}$, $\mathbf{I}^{NE}$, *are uniquely determined; further, prices for active firms,* p_j^{NE}, $j = 1, \ldots, n^*$, *are uniquely determined as well. All active firms make positive profits. Finally, the mass of consumers served in the NE is less than the socially optimal level, that is,* $Q^{NE} < Q^S$.

6.2. Uniqueness: Homogeneous Firms and Nonincreasing Returns to Investment

In the previous section we proved a uniqueness result under the assumption of constant returns to investment and log-concave demand. In this section we generalize our uniqueness result in one dimension, assuming nonincreasing returns to investment and a general demand function, but restrict it in another by assuming homogeneous firms. In particular, we establish uniqueness and symmetry of the NE in a model with homogeneous firms (if a NE exists).

First, we define the *marginal rate of substitution* (MRS) of I for x as the amount by which investment must increase per unit increase in demand quantity if the congestion cost level is to remain unchanged:

$$\mathrm{MRS}(I; x) = -\frac{\partial l(x, I)/\partial x}{\partial l(x, I)/\partial I}.$$

We introduce the following condition:

$$\frac{\partial}{\partial I}\mathrm{MRS}(x; I) \geqslant \frac{1}{x}, \quad \forall\, x > 0,\ I > 0, \tag{11}$$

Informallly, this condition ensures that the efficient investment level does not grow too rapidly as x increases; see Appendix EC.5 for details.

We have the following result.

Theorem 2. *Suppose Assumptions* 4 *and* 6 *hold. Suppose in addition that condition* (11) *holds. Suppose that the vectors of prices* $\mathbf{p}^{NE}$, *investment levels* $\mathbf{I}^{NE}$, *and demand levels* $\mathbf{x}^{NE}$ *form an NE; let* $Q^{NE} = \sum_{j=1}^{N} x_j^{NE}$.

Then $(\mathbf{p}^{NE}, \mathbf{I}^{NE})$ *is uniquely determined and symmetric. For all firms j, the NE demand quantities and investment levels are given by* $x_j^{NE} = Q^{NE}/N$ *and* $I_j^{NE} = I(Q^{NE}/N)$, *whereas prices are given by* (9). *All firms' profits are positive. Finally, the mass of consumers served in NE is less than the socially optimal level; that is,* $Q^{NE} < Q^S$.

Under Assumption 7, *the same result holds, except that* $Q^{NE} = Q^S = M$ *and prices are given by* (10). *In this case, the unique and symmetric NE is efficient.*

As a corollary, observe that for all congestion cost models studied in Lemma EC.1 in Appendix EC.1, the conclusion of Theorem 2 holds: these models satisfy Assumption 4 and can be shown to satisfy condition (11) (see Corollary EC.1 in Appendix EC.5).

6.3. Discussion

Theorems 1 and 2 establish a sharp prediction of firm decisions and consumer behavior in equilibirum; the NE is unique if it exists. Moreover, if firms are homogeneous the unique NE is symmetric. Our results are valid for the class of congestion models discussed in Lemma EC.1. Of particular importance is the fact that we do not assume convexity of the congestion cost function l, a common assumption made in papers that study these models (e.g., Xiao et al. 2007, Acemoglu and Ozdaglar 2007, Ozdaglar 2008, Hayrapetyan et al. 2007, Engel et al. 2004). Many of the congestion models discussed in Lemma EC.1, such as loss probabilities in queueing systems, do not satisfy this assumption. For example, Erlang's formula, $\mathrm{Erl}(x, I; s)$, is generally not jointly convex in x and I; it is not even convex in x for fixed I. However, as discussed after Lemma EC.1, $x\,\mathrm{Erl}(x, I; s)$ is jointly convex in x and I, and hence convex in x for fixed I.

It is also worth noting that a similar argument to the proof of Theorem 2 is valid for a pricing game without investment.[9] This result is of interest in itself; in particular, it provides the first uniqueness theorem for pricing games of this form in the literature.[10]

Theorem 2 suggests that for a broad class of models with homogeneous firms for which congestion cost exhibits nonincreasing returns to investment and firms choose prices and investments levels simultaneously, competition yields outcomes that are socially desirable. If demand is perfectly inelastic, the unique NE is efficient. If demand is not perfectly inelastic, there is an efficiency loss in the unique NE because the total mass of consumers served is less than socially efficient.[11] However, given the mass of consumers served in equilibrium, demand allocations and investment levels are efficient. We emphasize that even though firms are ex ante identical, they could be differentiated ex post by choosing different investment levels; this is not observed in equilibrium. The unique equilibrium is symmetric, and no dominant firm emerges. Note that uniqueness and symmetry is obtained even in models that exhibit constant returns to investment.

We conclude by discussing efficiency losses with heterogeneous firms, based on Theorem 1. When firms have different cost structures, the efficiency loss compared to the symmetric NE with homogeneous firms generally increases because firms with cost advantages can exploit their market power. In particular, the analysis in Theorem 1 explicitly considers participation, which is a fundamental determinant of market outcomes. Indeed, often more cost-efficient firms can price less efficient firms out of the market. We provide an example that illustrates these effects.

Example 1. We consider a duopoly ($N = 2$), in which $l_j(x_j, I_j) = g(\mathrm{Erl}(x_j, I_j; s_j))$, where $g(z) = z/(1 - z)$ and $\mathrm{Erl}(x_j, I_j; s_j)$ is Erlang's formula where x_j is the arrival rate, I_j is the service rate that is controlled by investment, and s_j is a predetermined number of servers (see

Appendix EC.1). We assume a linear demand function $P(q) = 3 - q$.

In the following analysis, we assume firm 1 has the most efficient technology with $s_1 = 3$ and corresponding cost coefficient $\xi_1 = 0.77$ (see Lemma 1), and we vary the cost efficiency of firm 2. We consider three cases, where firm 2 has three servers, two servers, or one server; the corresponding cost coefficients ξ_2 are 0.77, 1.1, and 2, respectively. In the social optimum firm 1 serves all demand Q^S, with $P(Q^S) = \xi_1$. In this case, $Q^S = x_1^S = 2.23$, and in the socially optimal solution total social surplus is equal to 2.5.

First, we consider a game where firms are homogeneous, so that both firms have $s_1 = s_2 = 3$. By Theorem 2 the NE is unique and symmetric if it exists. Indeed, the NE conditions yield $x_1^{\mathrm{NE}} = x_2^{\mathrm{NE}} = 0.95$ (see Proposition 1 and Lemma EC.4) and the associated social surplus only exhibits a 2.1% efficiency loss compared to the social optimum.

Second, we consider a game in which firm 2 has $s_2 = 2$, so firm 1 has a cost advantage. By Theorem 1 the NE is unique if it exists. Moreover, the NE conditions (see Proposition 2 and Lemma EC.4) yield $x_1^{\mathrm{NE}} = 1.07$, $x_2^{\mathrm{NE}} = 0.62$. The efficiency loss increases to 14.7%.

Third, we consider a game in which firm 2 has $s_2 = 1$; in this case firm 1 has an even larger cost advantage. The NE conditions yield $x_1^{\mathrm{NE}} = 1.12$, $x_2^{\mathrm{NE}} = 0$. Firm 1 can take advantage of its technological advantage due to a larger capacity, and price firm 2 out of the market. In this case, the efficiency loss increases significantly to 25%.

The example suggests that with asymmetry, inefficiency appears to increase. The main reason the inefficiency in the second game is larger than in the game with homogeneous firms is that in the second game, firm 2 serves consumers despite having an inferior technology. It is worth noting that in the third game the efficient firm serves all consumers, yet it realizes an inefficient operating point. This is because firm 1 exploits its cost advantage to extract a strong (and inefficient) markup, and thus the mass of consumers it serves in equilibrium is significantly less than the efficient level.

7. Existence of Nash Equilibrium

In this section we study conditions under which pure-strategy Nash equilibria exist for the pricing and investment game under consideration. In general, an NE may not exist, and we provide such an example. However, we provide several sufficient conditions that guarantee existence of an NE. In §7.1, we discuss two general existence theorems. We first show that if demand is concave, and all firms' congestion cost functions are concave as a function of demand, then an NE always exists. We then find a sufficient condition for existence of NE when firms exhibit constant returns to investment; notably, here we do not require the congestion cost function to be concave.

We conclude in §7.2 by discussing existence theorems that assume firms are homogeneous. In this case, we obtain two existence theorems: The first requires that firms exhibit constant returns to investment but face an elastic demand curve; the second assumes a perfectly inelastic demand curve but only requires nonincreasing returns to investment. In both cases, these existence results reveal that an NE exists only if there are sufficiently many competing firms relative to the elasticity of the congestion function.

We first show that even under the assumptions of Theorem 2, an NE need not exist. Similar examples can be constructed under the assumptions of Theorem 1.

Example 2. Consider a duopoly with homogeneous firms ($N = 2$) that face a perfectly inelastic demand of size $M = 10$. The congestion cost is $l(x, I) = x^6/I$. It is easy to see that the assumptions of Theorem 2 hold; if an NE exists, it must be unique and symmetric. By Theorem 2, the candidate NE price and investment level are given by (10) and $I_j^{\mathrm{NE}} = I(M/N)$, which lead to $p^{\mathrm{NE}} = 671$ and $I^{\mathrm{NE}} = 280$. Given that firm 2 is pricing at 671 and investing 280, the best response of firm 1 is to price at 1,367 and invest 20.5 to obtain a demand of 2.37 and a profit of 3,222. If firm 1 were to price at 671 and invest 280, it would only obtain a profit of 3,075. Thus, in this setting, firm 1 is better off investing less, attracting fewer consumers, and pricing higher than suggested by the expressions associated with the symmetric NE.

The preceding example suggests that if congestion cost increases too quickly as demand increases, there may be no NE. Motivated by this fact, in the next two sections we provide results that provide sufficient conditions for existence of an NE.

7.1. Existence: Heterogeneous Firms

Our first existence theorem requires that the congestion cost function and the inverse demand function are both concave in quantity. All proofs for this section can be found in Appendix EC.6.

Theorem 3. *Suppose Assumption* 4 *holds. Suppose in addition that for all j, $l_j(x, I)$ is a concave function of x for all $I > 0$, and that the inverse demand function $P(q)$ is a concave function of q where it is positive. Then there exists an NE.*

Note that concavity of the inverse demand function $P(q)$ in q is equivalent to concavity of the demand function $D(\Delta)$ in Δ. We also observe that, for example, concavity of the congestion cost function is satisfied by Erlang's formula for an $M/G/1/1$ queueing system. At the same time, it is a restrictive assumption that we alleviate in our next existence result.[12]

In our next result, we assume that firms exhibit constant returns to investment.

Proposition 5. *Suppose Assumption* 5 *holds. Furthermore, assume that for each firm j, ξ_j, and ϕ_j are defined as in Lemma* 1. *Suppose the demand function $D(\Delta)$ is a concave function of Δ where it is positive.*

Let $\bar{\xi} = \max_i \xi_i$, and define $A(\Delta)$ and $B(\Delta)$ for $\Delta > \bar{\xi}$ as follows:

$$A(\Delta) = \frac{\sum_i \Gamma_i(\Delta) - N + 1}{\sum_i (1 - \Gamma_i(\Delta))/\phi_i}; \quad B(\Delta) = -\frac{D'(\Delta)}{D(\Delta)},$$

where $\Gamma_i(\Delta) = [\phi_i(\Delta - h_i(\phi_i))]^{-1}$. Assume that $\lim_{\Delta \to \bar{\xi}} A(\Delta) > \lim_{\Delta \to \bar{\xi}} B(\Delta)$.

Finally, let $\bar{\Delta} = \sup\{\Delta\colon D(\Delta) > 0\}$, and assume for each j that $\bar{\Delta} + h_j(y_j)$ is a log-concave function of y_j and that $\bar{\xi} < \bar{\Delta}$. Then there exists an NE in which all firms are active.

This proposition is proven by using the approach of Theorem 1 to find a candidate NE where each firm's profit is locally stationary; we then apply the assumptions to show that at this candidate NE, each firm's best-response problem is concave, and so no firm has any incentive to deviate. Although the preceding proposition is theoretically appealing, the conditions over $A(\Delta)$ and $B(\Delta)$ do not directly provide insight into the structure of equilibrium; further, the condition that $\bar{\Delta} + h_j(y_j)$ must be log concave, although weaker than concavity of $h_j(y_j)$, can in fact be quite strong. In the next section, where we consider firms that are homogeneous, we obtain a related result that provides greater insight into conditions under which equilibria will exist.

7.2. Existence: Homogeneous Firms

In this section we suppose throughout that Assumption 6 holds. We start by considering a version of the existence result in Proposition 5, but with homogeneous firms.[13]

Theorem 4. *Suppose Assumptions* 5 *and* 6 *hold. In addition, suppose that the function $h(x)$ is log concave and that the demand function $D(\Delta)$ is a concave function of Δ where it is positive. Suppose also that the following inequality holds*:

$$N \geqslant \frac{(1 + e_h)^2}{1 + e_h(1 + e_d)}, \tag{12}$$

where $e_h = \phi h'(\phi)/h(\phi)$, $e_D = -\xi D'(\xi)/D(\xi)$, and ϕ and ξ are defined as in Lemma 1. *Then there exists an NE.*

The preceding condition directly relates the elasticity of the congestion cost function, the elasticity of the demand function, and the number of firms together to provide a condition for existence of equilibrium. In particular, note that a higher demand elasticity makes the condition less restrictive. Also, note that for large enough N the condition above is satisfied. Furthermore, note that $e_D \geqslant 0$; thus, a sufficient condition for existence of NE is:

$$N \geqslant e_h + 1. \tag{13}$$

This condition suggests that for existence of NE, the number of firms should be sufficiently large relative to the elasticity of the congestion cost function.

We conclude with a second result that assumes homogeneous firms. In this result we require stronger assumptions about demand—we assume it is perfectly inelastic—but no longer assume the congestion cost exhibits constant returns to investment; however, we do require the congestion cost function itself to be convex. Because demand is inelastic, we obtain an analog of (13) as a sufficient condition for existence of equilibrium.

Theorem 5. *Suppose Assumptions* 4, 6, *and* 7 *hold. In addition, suppose for all $I > 0$ that $l(x, I)$ is convex in x, and $(\partial l(x, I)/\partial x)/l(x, I)$ is nonincreasing in x. Suppose also that the number of firms N satisfies*:

$$N \geqslant \frac{x \partial l(x, I)/\partial x}{l(x, I)} + 1$$

for $x = M/N$ and $I = I(M/N)$. Then there exists an NE.

To construct concrete examples of the preceding results, consider congestion cost functions of the form $l(x, I) = (x/I)^q$, $q \geqslant 1$; this is a loss model derived from the exceedance probability of an $M/M/1$ queue (see Appendix EC.1). If $N \geqslant q + 1$, then an equilibrium is guaranteed to exist. The latter example clearly shows that to guarantee existence of NE, the congestion cost function cannot be too steep as demand increases, in relation to the number of incumbent firms. Indeed, observe that $l(x, I) = (x/I)^q$ satisfies the assumptions of either Theorem 4 or Theorem 5. In fact, if the steepness condition is not satisfied, a firm's best-response problem may fail to be concave, or even quasi-concave. Hence, necessary optimality conditions are not sufficient. As illustrated in Example 2, in these cases a firm may be better off by investing less and attracting fewer consumers than suggested by the symmetric NE. If the number of firms is small, the symmetric NE allocates a relatively large mass of consumers to each firm. In addition, if the congestion cost is steep (relative to the number of firms), firms will be heavily congested. By serving fewer consumers, a firm may significantly decrease its congestion level and as a consequence increase prices and profits.

7.3. Discussion

In this section we have provided existence results; these naturally complement our uniqueness results. In particular, if the assumptions of either of the preceding results holds together with the uniqueness conditions in Theorems 1 and 2, then *a unique NE exists.*

We note that the pricing and investment game is generally neither concave nor supermodular, so standard existence arguments do not apply (Vives 2001). The fact that NE may fail to exist is not entirely surprising if one considers that in Edgeworth-Bertrand competition, pure-strategy Nash equilibria may not exist (Edgeworth 1925, Levitan and Shubik 1972, Kreps and Scheinkman 1983, Tirole 1988). However, in Edgeworth-Bertrand games where firms compete by setting quantities and prices *simultaneously*, pure-strategy Nash equilibria generally do not exist, in marked contrast to our result (Acemoglu et al. 2009, Levitan and Shubik 1978).

As far as we know, Theorems 3, 4, and 5 are the first results in the literature concerning existence of a pure-strategy Nash equilibrium for congestion games where firms compete by simultaneously setting prices and investment levels. Acemoglu and Ozdaglar (2007), Baake and Mitusch (2007), and Engel et al. (2004) provide conditions for existence of pure-strategy Nash equilibrium for congestion games where firms only compete in prices, but not in investment. Their conditions have a similar spirit to ours; they restrict the "steepness" of the congestion cost functions. Note that of course, Theorems 3, 4, and 5 are also valid for a pricing game without investment (i.e., where $l_j(x, I)$ does not depend on I).

Mendelson and Shneorson (2003) provide an existence result where firms compete by simultaneously choosing investments and the mass of consumers served (as opposed to prices). Finally, Allon and Federgruen (2008, 2007), Cachon and Harker (2002), and So (2000) study existence of pure-strategy Nash equilibria in games where firms compete by choosing prices and "service levels." In the context of our model, this would imply that a firm commits to a fixed level of congestion ex ante, and implicitly agrees to invest as necessary to meet the service level. By contrast, in our model, a firm commits ex ante to investment expenditures instead.

8. Entry

Thus far, we have analyzed models given the existence of N incumbent firms that have already entered the market. In this section we study the efficiency properties of *entry* decisions made by profit-maximizing firms. We show that for a wide range of models, generally the free entry equilibrium number of firms may exceed the level that a social planner would choose; however, the free entry equilibrium becomes asymptotically efficient as the fixed cost of entry decreases to zero.

We assume that there exists an infinite number of homogeneous firms, and that any firm that wishes to enter the market must pay a strictly positive fixed sunk entry cost F to participate. To further simplify the analysis, in this section, we assume constant returns to investment. *Hence, throughout this section we assume Assumptions* 5 *and* 6 *hold.* (In Appendix EC.8 we present analogous results that assume inelastic demand, but allow us to consider models with nonincreasing returns to investment.)

First, we introduce a game-theoretic model to analyze competition between profit-maximizing firms. We consider the following two-stage game. In the first stage, firms simultaneously decide whether to enter and participate in the industry. In the second stage, incumbent firms compete by simultaneously setting prices and investment levels as described in §5. A *free entry equilibrium* is a pure-strategy subgame-perfect equilibrium of the two-stage game.

We showed in §§6 and 7 that for a wide range of models with homogeneous firms that exhibit constant returns to investment, a unique and symmetric NE exists. In this section we restrict attention to models that exhibit this type of postentry behavior.

Assumption 8. *In the postentry game where incumbent firms choose prices and investment levels simultaneously, for all numbers of incumbent firms, there exists a unique and symmetric NE.*

In light of the preceding assumption, it is useful to explicitly define profits in a symmetric NE as a function of the number of incumbent firms. Given N, let $\Pi(N)$ denote the profit an incumbent firm garners in a symmetric NE. Note that $\Pi(N) = P(Q_N)q_N - v(q_n) - F$, where q_N is the mass of consumers served by a firm in the postentry symmetric NE when there are N incumbent firms, and $Q_N = Nq_N$; see Proposition 2. The following definition formalizes the notion of a free entry equilibrium.

Definition 10. A *free entry equilibrium number of firms*, $N^E \in \{1, 2, \ldots\}$, satisfies $\Pi(N) \geqslant 0$ and $\Pi(N+1) < 0$.[14]

We make the following standard assumption.

Assumption 9. $\Pi(1) \geqslant 0$.

A key insight in our analysis is to observe that $\Pi(N)$ is similar to the profit obtained in a standard oligopoly postentry model with cost function $v(q) = \xi q$. As a consequence, we apply the results of Mankiw and Whinston (1986), which characterize entry in this setting. Following their approach, to compare against equilibrium outcomes, we consider as a benchmark the *second-best* problem faced by a social planner that chooses the number of participant firms in the industry, but that is unable to control the postentry behavior of firms; we assume that firms behave according to NE in the second stage. This is in contrast to §3. We introduce the following definition.

Definition 11. A number of firms N^S is *socially optimal* if it maximizes total social surplus assuming that firms play the unique (symmetric) NE strategy in the second stage, i.e., if it solves:[15]

$$\text{maximize } W(N, F) \equiv \int_0^{Q_N} P(q)\,dq - Nv(q_N) - NF. \qquad (14)$$

In our first result, we compare the free entry equilibrium with the socially optimal number of firms and show that, in general, there is excessive free entry. Indeed, Mankiw and Whinston (1986) show that in oligopoly models with increasing and convex cost functions, and downward-sloping demand function, under the following three assumptions excessive entry is obtained:

1. Q_N is strictly increasing in N and $\lim_{N\to\infty} Q_N = \bar{Q} < \infty$.
2. q_N is strictly decreasing in N.
3. $P(Q_N) - v'(q_N) \geqslant 0$, $\forall N$.

We have the following result. The proof relies on showing that conditions (1), (2), and (3) above hold in our model under the assumptions stated in the theorem; full details are provided in Appendix EC.7.

Theorem 6. *Suppose Assumptions* 5, 6, 8, *and* 9 *hold and that the inverse demand function $P(q)$ is a concave function of q. Then the free entry equilibrium number of firms exists and is unique, and it is no smaller than one less than any socially optimal number of firms.*

The thrust of the result is that, in general, there is more entry than the socially efficient level. An additional entrant creates social surplus equal to its profits. On the other hand, it generates a "business stealing effect:" an additional entrant marginally reduces the mass of consumers served by each of its competitors (condition (2) above). The business-stealing effect is not internalized by the additional entrant, generating more entry than is socially optimal.

Theorem 6 reveals excessive entry in models with strictly positive sunk entry costs. In the next result, we show that entry becomes asymptotically efficient as the sunk entry cost becomes small. Let $N^E(F)$ and $N^S(F)$ be the free entry and socially optimal number of firms, respectively, when the sunk entry cost is F. Mankiw and Whinston (1986) show that if conditions (1)–(3) above together with the condition $\lim_{N\to\infty} P(Q_N) - v'(q_N) = 0$ are satisfied, then entry becomes asymptotically efficient as $F \to 0$. The congestion cost function, l, and the demand function, P, remain the same for all sunk entry costs. Under the assumptions in Theorem 6, the latter condition is also satisfied, and, hence, we obtain the following result. The proof is direct from the proof of Theorem 6 and is omitted.

Theorem 7. *Suppose Assumptions* 5, 6, 8, *and* 9 *hold and that the inverse demand function $P(q)$ is a concave function of q. Then,* $\lim_{F\to 0} N^E(F) = \infty$ *and* $\lim_{F\to 0} N^S(F) = \infty$, *and* $\lim_{F\to 0} W(N^S(F), F) - W(N^E(F), F) = 0$.

Note that if firms were "price-takers" and, hence, the symmetric NE price was equal to the Pigovian price, then the free entry equilibrium number of firms would be socially optimal. The result implies that as the sunk entry cost becomes small, the free entry equilibrium number of firms grows to infinity. As a consequence, firms indeed become "price-takers" and the free entry equilibrium number of firms becomes socially optimal asymptotically.

We note that the results by Mankiw and Whinston (1986) cannot be directly applied in a model with a perfectly inelastic demand function.[16] In Appendix EC.8 we study entry and prove similar results to those in this section for such a model. There, we more generally assume nonincreasing returns to investment.

9. Conclusion

Our paper analyzes a model of investment and market structure in industries with congestion effects. Our model and results provide a framework through which competition in a range of congested service industries can be studied, yielding insight into business and policy considerations, with a particular emphasis on technology-based services. Our analysis highlights several key industry features that must be taken into account when characterizing industry performance:

1. *Cost structure*. Not surprisingly, the structure of costs has a critical impact on market outcomes; whereas congestion cost functions derived from loss systems impose a form of nonincreasing returns to investment, congestion cost functions derived from delay models impose a form of increasing returns to investment. In the latter, the socially efficient outcome calls for a single operating firm and a natural monopoly arises. In the former, competition among homogeneous firms yields symmetric equilibria and no dominant firm emerges.

2. *Timing of decisions*. In our model we assume that investment and pricing occur on the same timescale. A natural alternative is to consider a two-stage game where investment decisions are made prior to pricing decisions. In this model it is as if investment decisions involve a longer-term commitment than price decisions. In this case, one can construct examples where, in marked contrast to the efficient investment (conditional on the mass of consumers served) observed in the NE of the simultaneous pricing and investment game with homogeneous firms, highly inefficient investments are obtained in equilibrium. Firms may underinvest in the first stage to "soften" price competition in the second stage (see also De Borger and Van Dender 2006).

3. *Contractual structure*. In our model, firms compete by setting prices and investment levels simultaneously. This represents a *best effort* (BE) contractual agreement, where firms provide the best possible service given their infrastructure, but without an explicit guarantee. For example, typical end-user Internet service provision contracts disclaim liability for loss or delay. A common alternative is a model where firms compete by setting prices and *service level guarantees* (SLGs) simultaneously. The SLG is a contractual obligation on the part of the service provider: regardless of how many customers subscribe, the firm is responsible for investing so that the congestion experienced by all subscribers is equal to the SLG. In some industries, service-level guarantees are the norm (e.g., expedited shipping, such as FedEx and UPS). DiPalantino et al. (2009) compare these competitive models and show that equilibria can be drastically different. For example, in the case of constant returns to investment and homogeneous firms, although the Nash equilibrium price for the SLG game is perfectly competitive, firms obtain positive markups in the unique Nash equilibrium for the BE game.

Our paper leaves many significant directions for future research. Our model has considered consumers that are homogeneous in their preferences: all consumers trade off congestion and money in the same way. We leave for future research study of a model where consumers have heterogeneous preferences. We have not modeled the fact that consumers may face switching costs in moving between providers. We have also not modeled the fact that firms may choose to contract with each other, particularly in providing services that exhibit strong network effects. Indeed, in such industries we might see integration across firms as well. We leave modeling of these additional phenomena to future work.

10. Electronic Companion

An electronic companion to this paper is available as part of the online version that can be found at http://or.journal.informs.org/.

Endnotes

1. Our results can be easily extended to a setting where all firms additionally face a constant cost per consumer served. However, to simplify the model and notation, we ignore this cost.

2. We assume throughout the paper that derivatives at zero are right-directional derivatives.

3. In §8 we consider the setting where a social planner chooses the number of firms.

4. For any $x_j > 0$, problem (5) admits an optimal solution because the objective function is continuous and coercive for $I_j > 0$. For $x_j = 0$, we define by convention that the optimal solution is $I_j = 0$.

5. Under our assumptions it can be shown that $v_i(x_i)$ is differentiable for $x_i > 0$.

6. Note that in the case of decreasing returns to investment, if a firm can costlessly divide itself into multiple facilities, it will always choose to do so; the resulting cost structure will exhibit constant returns to investment.

7. Throughout the paper, K_j being convex refers to K_j being convex on the set $[0, \infty) \times (0, \infty)$.

8. Campo-Rembado and Sundararajan (2004) use such a model to study competition among wireless service providers.

9. In that case, if $l(x)$ is nondecreasing and $xl(x)$ is a convex function of x, then, if an NE exists, it is unique and symmetric.

10. Baake and Mitusch (2007) and De Borger and Van Dender (2006) provide uniqueness results, but only for a duopoly.

11. That $Q^{\mathrm{NE}} < Q^S$ was also found by Xiao et al. (2007) in the particular case of industries that exhibit constant returns to investment, and by Ozdaglar (2008) and Engel et al. (2004) when firms only compete in prices.

12. In particular, if $l(x, I) = \mathrm{Erl}(x, I; s)$, $s > 1$, an NE may fail to exist.

13. In this case, the condition $\lim_{\Delta\to\bar{\xi}} A(\Delta) > \lim_{\Delta\to\bar{\xi}} B(\Delta)$ is satisfied directly. To prove the existence result below, we assume $h(x)$ log concave; under condition (12) we do not need to assume the stronger condition that $\Delta + h(x)$ is log concave.

14. If Assumption 8 holds, a free entry equilibrium number of firms is the number of entrants in a subgame-perfect equilibrium of the two-stage entry game. The first condition in the definition guarantees that entrants are better off participating in the industry. The second condition ensures that a potential additional entrant prefers not to enter.

15. Because $F > 0$, in any socially optimal solution, the number of entrants is finite.

16. The complication arises from the fact that the full price in the market cannot be expressed as $P(q)$.

Acknowledgments

The authors are grateful for helpful conversations with Daron Acemoglu, Gad Allon, Lanier Benkard, Eduardo Engel, Awi Federgruen, Sunil Kumar, Jon Levin, Asuman Ozdaglar, Tim Roughgarden, Ilya Segal, Arun Sundararajan, and Eva Tardos, as well as with seminar participants at Cornell University, INFORMS, Kellogg School of Management, Massachusetts Institute of Technology, and Stanford University. They also thank the area editor Izak Duenyas, the associate editor, and two anonymous reviewers for constructive comments and suggestions that greatly improved the paper. This work was supported in part by the National Science Foundation under grants 0428868 and IIS-0428868, and by the Okawa Foundation.

References

Acemoglu, D., A. Ozdaglar. 2007. Competition and efficiency in congested markets. *Math. Oper. Res.* **32**(1) 1–31.

Acemoglu, D., K. Bimpikis, A. Ozdaglar. 2009. Price and capacity competition. *Games Econom. Behav.* **66**(1) 1–26.

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

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

Baake, P., K. Mitusch. 2007. Competition with congestible networks. *J. Econom.* **91**(2) 151–176.

Beckmann, M., C. B. McGuire, C. B. Winsten. 1956. *Studies in the Economics of Transportation*. Yale University Press, New Haven, CT.

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

Campo-Rembado, M. A., A. Sundararajan. 2004. Competition in wireless telecommunications. Working paper, New York University, New York.

De Borger, B., K. Van Dender. 2006. Prices, capacities, and service levels in a congestible Bertrand duopoly. *J. Urban Econom.* **60** 264–283.

Deneckere, R., J. Peck. 1995. Competition over price and service rate when demand is stochastic: A strategic analysis. *RAND J. Econom.* **26**(1) 148–162.

De Vany, A. S., T. R. Saving. 1983. The economics of quality. *J. Political Econom.* **91**(6) 979–1000.

DiPalantino, D., R. Johari, G. Y. Weintraub. 2009. Competition and contracting in service industries. Working paper, Stanford University, Stanford, CA.

Edgeworth, F. 1925. The pure theory of monopoly. *Papers Relating to Political Economy*, Vol. 1. Macmillan, London, 111–142.

Engel, E. M., R. Fischer, A. Galetovic. 2004. Toll competition among congested roads. *Topics Econom. Anal. Policy* **4**(1).

Hall, J., E. Porteus. 2000. Customer service competition in capacitated systems. *Manufacturing Service Oper. Management* **2**(2) 144–165.

Hayrapetyan, A., E. Tardos, T. Wexler. 2007. A network pricing game for selfish traffic. *Distributed Comput.* **4**(12) 255–266.

Kleinrock, L. 1975. *Queueing Systems, Volume 1: Theory*, 1st ed. Wiley-Interscience, Malden, MA.

Kreps, D. M., J. A. Scheinkman. 1983. Quantity precommitment and Bertrand competition yield Cournot outcomes. *Bell J. Econom.* **14**(2) 326–337.

Levitan, R., M. Shubik. 1972. Price duopoly and capacity constraints. *Internat. Econom. Rev.* **13**(1) 111–122.

Levitan, R., M. Shubik. 1978. Duopoly with price and quantity as strategic variables. *Internat. J. Game Theory* **7**(1) 1–11.

Mankiw, N. G., M. D. Whinston. 1986. Free entry and social inefficiency. *RAND J. Econom.* **17**(1) 48–58.

Mendelson, H., S. Shneorson. 2003. Internet peering, capacity and pricing. Working paper, Stanford University, Stanford, CA.

Ozdaglar, A. 2008. Price competition with elastic traffic. *Networks* **52**(3) 141–155.

Pigou, A. C. 1920. *The Economics of Welfare*, 1st ed. Macmillan, London.

Roughgarden, T. 2005. *Selfish Routing and the Price of Anarchy*. MIT Press, Cambridge, MA.

Scotchmer, S. 1985. Profit-maximizing clubs. *J. Public Econom.* **27** 25–45.

Scotchmer, S. 2002. Local public goods and clubs. *Handbook of Public Economics*, Vol. 4. Elsevier, Amsterdam, 1997–2042.

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

Tirole, J. 1988. *The Theory of Industrial Organization*, 1st ed. MIT Press, Cambridge, MA.

Vives, X. 2001. *Oligopoly Pricing*, 1st ed. MIT Press, Cambridge, MA.

Wardrop, J. G. 1952. Some theoretical aspects of road traffic research. *Proc. Inst. Civil Engineers, Pt. II* **1** 325–378.

Xiao, F., H. Yang, D. Han. 2007. Competition and efficiency of private toll roads. *Transportation Res. Part B* **41** 292–308.