Submitted to *Operations Research*
manuscript (Please, provide the mansucript number!)

Investment and Market Structure in Industries with Congestion

First Version: January 2007; This Version: August, 2009

Ramesh Johari
Stanford University, Stanford, CA, 94305, ramesh.johari@stanford.edu

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

Benjamin Van Roy
Stanford University, Stanford, CA, 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, while 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.

Key words: Competition, game theory, services, congestion

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 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 oppose 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 generalization 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 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 vs. 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 as well as 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.

It is worth noting that although we assume convexity of the total congestion cost to obtain these results, we do not assume convexity of the *per user* congestion cost function—a common assumption made in papers that study related models (e.g., Acemoglu and Ozdaglar (2007), Xiao et al. (2007)). In particular, our model captures some important congestion models that exhibit nonincreasing returns to investment, but where the congestion cost function is not convex. An example is where cost corresponds to loss probability in a queueing system.

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 study the efficiency properties of entry decisions made by homogeneous profit-maximizing firms. In that analysis, we assume 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. (2008), Acemoglu and Ozdaglar (2007), Cachon and Harker (2002), Scotchmer (1985), and the detailed discussion in Section 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, as 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 (WiFi) 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 WiFi 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. Since 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. Further, 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." Further, 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 Section 2 we review literature related to our work. In Section 3 we introduce our model of service provision. In Section 4 we introduce and study the notion of returns to investment. In Section 5 we introduce a game theoretic model to analyze competition between profit-maximizing firms, and characterize its Nash equilibrium. In

Section 6 we study uniqueness of Nash equilibrium. In Section 7 we study existence of equilibria. In Section 8 we study entry decisions made by firms. Finally, in Section 9 we conclude and provide some thoughts for future research.

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. Further, 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. Johari and Weintraub (2008) 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, while 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) that independently and 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, while 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 pure strategy Nash equilibrium. Specific bounds are obtained for *symmetric* equilibria when firms are assumed to be homogeneous. 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. Further, 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. (2008) 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 pure strategy Nash equilibrium; Scotchmer (1985) does not study uniqueness and proves existence for large economies only. Vany and Saving (1983) analyze a similar model but considers 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 Section 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 $N \geq 2$ incumbent firms are present after entry decisions have been made; we consider a game with entry decisions in Section 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 post-entry 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* $\ell_j(x_j, I_j)$. The congestion cost function $\ell_j(x_j, I_j)$ represents the disutility perceived by consumers due to congestion.

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

[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.

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

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 $q \geq 0$, $P(q)$ is nonnegative, and continuously differentiable with $P'(q) < 0$ whenever positive. Further,* $\lim_{q\to\infty} P(q) = 0$.[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), and 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 $\boldsymbol{p}$ and $\boldsymbol{I}$, a vector of demand quantities $\boldsymbol{x} \geq 0$ is a *Wardrop equilibrium* if

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

$$p_j + \ell_j(x_j, I_j) \geq P(Q), \qquad \text{for all } j, \tag{3}$$

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

Under Assumptions 1 and 2, given price and investment vectors $\boldsymbol{p}$ and $\boldsymbol{I}$, 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 $W(\boldsymbol{p}, \boldsymbol{I})$. When $I_j = 0$ for all j, we let $W(\boldsymbol{p}, \boldsymbol{I}) = \emptyset$.

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 $N \geq 2$ of incumbent firms.[3]

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

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

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

$$\begin{aligned} &\text{maximize} \quad \int_0^{\sum_{i=1}^N x_i} P(q)dq - \sum_{i=1}^N (x_i \ell_i(x_i, I_i) + I_i) \\ &\text{subject to} \quad \boldsymbol{x}, \boldsymbol{I} \geq 0. \end{aligned} \tag{4}$$

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 $x_j \geq 0$, 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]

$$v_j(x_j) \equiv \min_{I_j \geq 0} \quad [x_j \ell_j(x_j, I_j) + I_j]. \tag{5}$$

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

Observe that since 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 $\boldsymbol{x} \geq 0$, 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* which yields a unifying framework for classifying cost structures arising from different congestion models. We start with the following definition; we use $\ell(x, I)$ to generically refer to the congestion cost function of a given firm.

[4] For any $x_j > 0$, problem (5) admits an optimal solution since the objective function is continuous and coercive for $I_j > 0$. For $x_j = 0$, the optimal solution is $I_j = 0$.

[5] Under our assumptions it can be shown that $v_i(x_i)$ is differentiable for $x_i \geq 0$.

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

Recall that $\ell(x,I)$ 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 $\ell(x,I)$ exhibits *nonincreasing (resp., nondecreasing) returns to investment* if:

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

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 $K_j(x_j, I_j) = x_j \ell_j(x_j, I_j)$ is jointly convex in (x_j, I_j) and strictly convex in I_j.*[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 4 in Appendix B. Finally, the optimal solution of problem (5) is also unique in this case; that is, for all j and $x_j \geq 0$, 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.

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

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 $\ell_j(x,I) = h_j(x/I)$.*

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 $\ell(x,I) = h(x/I)$. Then there exists a unique solution ϕ to the equation $\phi^2 h'(\phi) = 1$. Further, $I(x) = x/\phi$, and thus $v(x) = \xi x$, where:*

$$\xi = h(\phi) + \frac{1}{\phi} > 0.$$

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 $\ell(q,I)$ 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 A 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 A, an important class of congestion models with this property are derived from steady-state expected waiting times in queueing models.

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

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 A, industries where consumers are *loss sensitive* exhibit *nonincreasing returns* to investment, while 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 post-entry 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 Sections 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 $\boldsymbol{p}^{NE}$, investment levels $\boldsymbol{I}^{NE}$, and demand quantities $\boldsymbol{x}^{NE}$, 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: $\boldsymbol{x}^{NE} \in W(\boldsymbol{p}^{NE}, \boldsymbol{I}^{NE})$.
2. Each firm maximizes profit given prices and investment levels of other firms; i.e., for all $j = 1, \ldots, N$, $p_j, I_j \geq 0$, and $\boldsymbol{x} \in W(p_j, \boldsymbol{p}_{-j}^{NE}, I_j, \boldsymbol{I}_{-j}^{NE})$,

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

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

In Section 5.1, we provide Nash equilibrium conditions for the general model. In Section 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 $(\boldsymbol{p}^{NE}, \boldsymbol{I}^{NE}, \boldsymbol{x}^{NE})$ if it invests a positive amount $I_j^{NE} > 0$.

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 C.

PROPOSITION 1. *Suppose that the vectors of prices $\boldsymbol{p}^{NE}$, investment levels $\boldsymbol{I}^{NE}$, and demand levels $\boldsymbol{x}^{NE}$ form a 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:*

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

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

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

In the NE, firm $j \in A$ makes a profit equal to $\pi(p_j^{NE}, I_j^{NE}, x_j^{NE}) = P(Q^{NE})x_j^{NE} - v_j(x_j^{NE})$. Further, 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 $(\boldsymbol{x}^S, \boldsymbol{I}^S)$, she should charge a *Pigovian price* for the service of each firm j, given by $p_j = x_j^S \partial \ell_j(x_j^S, I_j^S)/\partial x_j$ (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 $\ell_i(x,I) = \ell_j(x,I)$.*

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

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

DEFINITION 9. A NE is *symmetric* if $p_i^{NE} = p_j^{NE}$ and $I_i^{NE} = I_j^{NE}$ for all i, j; since the WE is uniquely defined, this also implies $x_i^{NE} = x_j^{NE}$. A NE is *symmetric among active firms* if $p_i^{NE} = p_j^{NE}$ and $I_i^{NE} = I_j^{NE}$ 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:

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

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 D, 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:

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

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 Section 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 Section 6.2 we assume firms are homogeneous, and consider a class of models that exhibit nonincreasing returns to investment; we show that if a 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 Section 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* $\xi_1 \leq \cdots \leq \xi_N$. That is, firm 1 has the lowest total cost function and firm N has the highest. Recall that in Appendix A we discuss important congestion cost functions that exhibit constant returns to investment.

The following result shows that a 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 E.

Proposition 2. *Suppose Assumption 5 holds. Suppose that the vectors of prices* $\boldsymbol{p}^{NE}$*, investment levels* $\boldsymbol{I}^{NE}$*, and demand levels* $\boldsymbol{x}^{NE}$ *form a 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* $n^* \in \{1, ..., N\}$ *such that all firms* $i \leq n^*$ *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* $n^* \geq 1$*. Suppose that the vectors of prices* $\boldsymbol{p}^{NE}$*, investment levels*

$\boldsymbol{I}^{NE}$, and demand levels $\boldsymbol{x}^{NE}$ form a NE with n^* active firms. Then, $\boldsymbol{x}^{NE}$ and $\boldsymbol{I}^{NE}$ are uniquely determined; further, prices for active firms, p_j^{NE}, $j = 1, ..., n^*$, 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 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 $\boldsymbol{p}^{NE}$, investment levels $\boldsymbol{I}^{NE}$, and demand levels $\boldsymbol{x}^{NE}$ form a NE with n^* active firms. Then, $n^* \geq 1$ and n^* is uniquely determined.*

The previous results directly lead 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 $\boldsymbol{p}^{NE}$, investment levels $\boldsymbol{I}^{NE}$, and demand levels $\boldsymbol{x}^{NE}$ form a NE; let n^* be the number of active firms, and let $Q^{NE} = \sum_{j=1}^{N} x_j^{NE}$.*

Then $n^* \geq 1$, and n^*, $\boldsymbol{x}^{NE}$, $\boldsymbol{I}^{NE}$, are uniquely determined; further, prices for active firms, p_j^{NE}, $j = 1, ..., n^*$, 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 key results that guarantee uniqueness and symmetry of the NE in a model with homogeneous firms (if a NE exists).

Recall that in a NE, an active firm j invests efficiently: $I_j^{NE} = I(x_j^{NE})$. We start with the following result.

Proposition 5. *Suppose Assumptions 4 and 6 hold. Suppose further that the effective congestion cost function $\ell(x, I(x))$ is nondecreasing for all $x > 0$. Then any NE must be symmetric among active firms.*

The preceding proposition is shown by recognizing that the NE conditions for the simultaneous pricing and investment game can be related to the NE conditions of a pricing game *without investment, but with an effective congestion cost function $\ell(x, I(x))$*. We show that in a game without investment, the NE is symmetric among active firms as long as the congestion cost function is increasing; the preceding proposition effectively enforces a similar condition in the game with investment: that $\ell(x, I(x))$ must be nondecreasing.

Interpreted informally, the preceding condition implies that the efficient investment level $I(x)$ cannot grow too rapidly as x increases. However, the condition that $\ell(x, I(x))$ is nondecreasing is not a condition over model primitives, so we provide the following characterization. Define the *marginal rate of substitution* 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:

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

In Appendix E (Lemma 6), we show that if Assumption 4 holds, and if for all $x > 0$ and $I > 0$ there holds:

$$\frac{\partial}{\partial I} \text{MRS}(x; I) \geq \frac{1}{x}, \tag{11}$$

then the effective congestion cost function $\ell(x, I(x))$ is nondecreasing for all $x > 0$.

This result is proven by considering the behavior of the efficient investment level as x increases. Suppose that the demand quantity x increases to $x + \epsilon$. One possibility is to invest sufficiently to *maintain* the same congestion level, $\ell(x, I(x))$. Under our assumption on the marginal rate of substitution, however, we show that the efficient choice of investment level allows the congestion level to *increase* from $\ell(x, I(x))$, so that corresponding savings are garnered on the investment cost. For this result to hold, the marginal rate of substitution of I for x must increase sufficiently quickly.

Finally, to obtain a uniqueness result, we also establish that in any NE *all* firms are active. The following result demonstrates that convexity of K suffices to ensure every firm participates.

Proposition 6. *Suppose Assumption 4 and 6 hold. Then, in any NE that is symmetric among active firms, all firms must be active, and all firms make positive profits.*

The following theorem is the main result of this section.

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

Then $(\boldsymbol{p}^{NE}, \boldsymbol{I}^{NE})$ 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 $I_j^{NE} = I(Q^{NE}/N)$, while 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 2 in Appendix A, the conclusion of Theorem 2 holds: these models satisfy Assumption 4, and can be shown to satisfy condition (11) (see Corollary 1 in Appendix E).

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 2. 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 2, 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 2, $x\,\mathrm{Erl}(x, I; s)$ 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

[9] In that case, if $\ell(x)$ is nondecreasing and $x\ell(x)$, is a convex function of x, then, if a 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 $Q^{NE} < Q^S$ 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.

equilibrium is symmetric and no dominant firm emerges. Note that uniqueness and symmetry is obtained even in models that exhibit constant returns to investment.

Xiao et al. (2007) derive bounds for the efficiency loss of symmetric NE in models with homogeneous firms. They found that the efficiency loss of symmetric NE compared to the social optimum is no more than 15% for exponential or linear demand functions. This result complements the discussion in the previous paragraph; efficiency loss of symmetric NE is not significant. Further, our uniqueness result strengthens the results in Xiao et al. (2007); it implies that their bounds are valid even if one considers asymmetric equilibria.

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. Xiao et al. (2007) makes a similar observation; however, in their analysis they assume constant returns to investment, and they also assume that in the socially optimal solution all firms participate. This is only possible if all firms share the *same* total cost coefficient $\xi_i = \xi$. In general, with heterogeneous firms, only the firms that have the *lowest* total cost coefficient $\xi_i = \min_j \xi_j$ will be active in the socially optimal solution.

Instead, our analysis is more general because we explicitly characterize the equilibrium (if it exists) with heterogeneous firms. In the proof of Theorem 1, we derive a threshold condition that identifies the active firms in an equilibrium; we can do this because we explicitly consider 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 $\ell_j(x_j, I_j) = g(\mathrm{Erl}(x_j, I_j; s_j))$, 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 A). 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 $x_1^{NE} = x_2^{NE} = 0.95$ (see Proposition 1 and Lemma 5) 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 5) yield $x_1^{NE} = 1.07,\ x_2^{NE} = 0.62$. 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 $x_1^{NE} = 1.12,\ x_2^{NE} = 0$. 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, a NE may not exist, and we provide such an example. However, we provide several sufficient conditions that guarantee existence of a NE. In Section 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 a 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 Section 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 a 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, a 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 faces a perfectly inelastic demand of size $M=10$. The congestion cost is $\ell(x,I)=x^6/I$. It is easy to see that the assumptions of Theorem 2 hold; if a NE exists, it must be unique and symmetric. By Theorem 2, the candidate NE price and investment level are given by (10) and $I_j^{NE}=I(M/N)$, which lead to $p^{NE}=671$ and $I^{NE}=280$. Given that firm 2 is pricing at 671 and investing 280, the best response of firm 1 is to price at 1367 and invest 20.5 to obtain a demand of 2.37 and a profit of 3222. If firm 1 were to price at 671 and invest 280, it would only obtain a profit of 3075. 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 a 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 F.

Theorem 3. *Suppose Assumption 4 holds. Suppose in addition that, for all j, $\ell_j(x,I)$ 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 a 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 firms exhibit constant returns to investment.

Proposition 7. *Suppose Assumption 5 holds. Further, 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 $\overline{\xi}=\max_i \xi_i$, and define $A(\Delta)$ and $B(\Delta)$ for $\Delta>\overline{\xi}$ 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 $\lim_{\Delta\to\overline{\xi}} A(\Delta)>\lim_{\Delta\to\overline{\xi}} B(\Delta)$.

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

Finally, let $\overline{\Delta} = \sup\{\Delta : D(\Delta) > 0\}$, and assume for each j that $\overline{\Delta} + h_j(y_j)$ is a log-concave function of y_j and that $\overline{\xi} < \overline{\Delta}$. Then there exists a 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. While 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 $\overline{\Delta} + h_j(y_j)$ must be log-concave, while 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 7, 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:*

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

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 a 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. Further, note that $e_D \geq 0$; thus a sufficient condition for existence of NE is:

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

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.

[13] In this case, the condition $\lim_{\Delta\to\overline{\xi}} A(\Delta) > \lim_{\Delta\to\overline{\xi}} B(\Delta)$ 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 $\overline{\Delta} + h(x)$ is log-concave.

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 $\ell(x,I)$ is convex in x, and $(\partial\ell(x,I)/\partial x)/\ell(x,I)$ is nonincreasing in x. Suppose also that the number of firms N satisfies:*

$$N \geq \frac{x\partial\ell(x,I)/\partial x}{\ell(x,I)} + 1$$

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

To construct concrete examples of the preceding results, consider congestion cost functions of the form $\ell(x,I) = (x/I)^q$, $q \geq 1$; this is a loss model derived from the exceedance probability of an M/M/1 queue (see Appendix A). If $N \geq q+1$, 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 $\ell(x,I) = (x/I)^q$ 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 quasiconcave. 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 less 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. 2008, Levitan and Shubik 1978).

As far as we know, Theorems 3, 4 and 5 are the first results in the literature concerning existence of 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 $\ell_j(x, I)$ does not depend on I).

Mendelson and Shneorson (2003) provide an existence result where firms compete by simultaneously choosing investments and the amount of consumers served (as opposed to prices). Finally, Allon and Federgruen (2008), Allon and Federgruen (2007), Cachon and Harker (2002), and So (2000) study existence of pure strategy Nash equilibrium 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.* (Due to space constraints, in Appendix H 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 Section 5. A *free entry equilibrium* is a pure strategy subgame perfect equilibrium of the two-stage game.

We showed in Sections 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 post-entry behavior.

Assumption 8. *In the post-entry 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 post-entry 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*, $N^E \in \{1, 2, ...\}$, satisfies $\Pi(N) \geq 0$ and $\Pi(N+1) < 0$.[14]

We make the following standard assumption.

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

A key insight in our analysis is to observe that $\Pi(N)$ is similar to the profit obtained in a standard oligopoly post-entry model with cost function $v(q) = \xi q$. As a consequence we apply the results of Mankiw and Whinston (1986) that 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 post-entry behavior of firms; we assume firms behave according to NE in the second stage. This is in contrast to Section 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]

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

[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] Since $F > 0$, in any socially optimal solution, the number of entrants is finite.

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

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 G.

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 H 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 have a critical impact on market outcomes; while 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 time scale. 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,

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

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). Johari and Weintraub (2008) 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, while 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.

Appendix A: Returns to Investment: Examples

In this appendix we provide a variety of examples that exhibit nonincreasing or constant returns to investment.

EXAMPLE 3. *Loss systems.* In this example we consider service provision or make-to-order facilities that can be modeled as loss systems (i.e., consumers leave if they find all servers busy) and where firms invest to increase the service rate. Formally, suppose that each firm owns a processing facility with s servers; that is, the firm can process at most s jobs simultaneously. The investment level I_j controls the service rate of each server; thus, if firm j invests I_j, then the service rate of each server is $\kappa(I_j)$. Suppose that processes arrive according to a Poisson process with rate x_j. If a process arrives to the facility and finds all servers busy, it is denied service and lost. Such a facility can be modeled as an $M/G/s/s$ queue. It is well known that the steady-state probability $\mathrm{Erl}(x_j, \kappa(I_j); s)$ that a job finds all servers busy and is not served is given by *Erlang's formula*:

$$\mathrm{Erl}(x_j, \kappa(I_j); s) \equiv \frac{\left(\frac{x_j}{\kappa(I_j)}\right)^s / s!}{\sum_{k=0}^{s} \left(\frac{x_j}{\kappa(I_j)}\right)^k / k!}.$$

Suppose consumers pay for subscription to a firm and this allows them to receive service whenever they request it, provided there are servers available. Additionally, suppose a consumer's disutility is a function of the average fraction of times his requests are not served because all servers are busy. Then a possible congestion cost function is given by $\ell(x,I) = g(\text{Erl}(x,\kappa(I);s))$, where g is a strictly increasing and convex function with $g(0) = 0$ and $g(1) = \infty$. The congestion cost increases with the expected fraction of times that consumers do not receive service when requesting it. One practical setting where such a model may apply arises in wireless service provision.[17] Alternatively, the expression for loss probability could be derived from an $M/G/1/s$ system or be modeled as the exceedance probability (the probability the queue exceeds a certain length) of an $M/M/1$ queue.[18]

Example 4. *Capacity sharing.* Suppose each firm owns a processing facility where investment expenditure determines processing capacity per unit time; i.e., if firm j invests I_j, it has a capacity of processing $\kappa(I_j)$ demand units per time unit. We assume κ is a concave function: the marginal productivity of investment expenditure in building capacity is nonincreasing. Hence, building a unit of capacity is not more expensive than building subsequent units. If x_j is the total mass of demand at firm j, and capacity is equally shared among the consumers, then each infinitesimal consumer faces a processing delay of $x_j/\kappa(I_j)$ time units. If consumers are most sensitive to delay, then $\ell(x_j,I_j) = x_j/\kappa(I_j)$. Alternatively, the marginal cost of delay perceived by consumers may be increasing; e.g., consumers may tolerate the first units of delay more than subsequent units. In this case, a more appropriate model is $\ell(x_j,I_j) = f(x_j/\kappa(I_j))$, where f is an increasing and convex function.

Example 5. *Capacitated systems with retrials.* Consider a simple model of a capacitated system with retrials. Suppose that at every period of time x consumers require service from a firm. The firm has a capacity to serve $I < x$ consumers simultaneously. If consumers are randomly sampled,

[17] In wireless services it is plausible 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. It is also reasonable to assume that consumers' marginal discontent increases as the fraction of unserved requests increases, eventually becoming unbounded as the fraction converges to one. Such a model can be captured by a congestion cost function of the form $g(\text{Erl}(x,\kappa(I);s))$. For example, suppose that firm j provides service to a population of x_j consumers that are uniformly distributed in a square area normalized to one. Firm j decides how many wireless base stations to build, $\kappa(I_j)$. We assume that base stations are located uniformly in the square and that consumers always connect to the closest base station. Therefore, the number of consumers served by each base station equals $x_j/\kappa(I_j)$. We suppose that given available spectrum bandwidth and transmission technology, firm j has s *effective channels* at each base station; i.e., each base station can serve the requests (or calls) of at most s consumers simultaneously. Each consumer generates requests according to a Poisson process with rate one, and each call has expected service time equal to one. In this setting, each base station can be modeled as an $M/G/s/s$ queue with load $x_j/\kappa(I_j)$. This model is similar to the one used by Campo-Rembado and Sundararajan (2004) to study competition in wireless telecommunications.

[18] Hall and Porteus (2000) use loss system models similar to those presented in the example to analyze competition in capacitated systems.

the probability that a given consumer gets service is equal to I/x. If a consumer does not receive service, she retries next period. The expected time until the request of a given consumer is met for the first time is given by $\frac{x}{I}$.

All the models introduced in the previous examples satisfy Assumptions 1 and 4, as we show in the following lemma. If $\kappa(I) = I$, they satisfy Assumption 5. An important step in proving the second part of the result is to use the fact that $x\,\text{Erl}(x, I; s)$ is jointly convex in x and I (Harel 1990).[19]

LEMMA 2. *Let $\kappa(I)$ be a strictly increasing, concave, and twice differentiable function for $I \geq 0$, with $\kappa(0) = 0$. Suppose that the congestion cost function $\ell(x, I)$, has one of the following forms:*

1. *$\ell(x, I) = f(x/\kappa(I))$, where f is strictly increasing, convex, and twice differentiable, with $f(0) = 0$ and $\lim_{z\to\infty} f(z) \to \infty$.*
2. *$\ell(x, I) = g(\text{Erl}(x, \kappa(I); s))$, where g is strictly increasing, convex, and twice differentiable, with $g(0) = 0$ and $g(1) = \infty$.*[20]
3. *$\ell(x, I) = f(x)/\kappa(I)$, where $f(x)/\kappa(I)$ is a jointly convex function of x and I.*

Then, the congestion cost, ℓ, satisfies Assumptions 1 and 4. Moreover, if $\kappa(I) = I$, then ℓ given by cases (1) and (2) above satisfies Assumption 5.

Proof. We prove the lemma in the case where $\kappa(I) = I$; the generalization for κ concave is straightforward, and we omit the details. In cases (1) and (2), it is straightforward to check that the congestion cost, ℓ, satisfies Assumption 1 and exhibits constant returns to investment. We show that Assumption 4 is satisfied.

Cases 1 and 2 of the main result follow by Lemma 3, given below. Case 1 follows by noting that $K(x, I) = x^2/I$ is a jointly convex function of x and I (since its Hessian is positive semidefinite), and strictly convex in I; and case 2 follows because $x\,\text{Erl}(x, I; s)$ is jointly convex in x and I, and strictly convex in I (Harel 1990).

For case 3, K is convex if and only if it has positive semidefinite Hessian; and this in turn is equivalent to (i) $K_{xx}(x, I) \geq 0$ and (ii) $\det[\text{Hess}\, K(x, I)] \geq 0$ (det denotes determinant and Hess denotes Hessian).[21] In case 3, $K_{xx}(x, I) = (2f'(x) + xf''(x))/I \geq 0$, because f is a strictly increasing and convex function. We have that

$$\det[\text{Hess}\, K(x, I)] = \frac{f(x)\,(xf'(x) - f(x)) + x\,(f(x)f'(x) + 2xf(x)f''(x) - xf'(x)^2)}{I^4}\,. \qquad (15)$$

[19] We note, however, that $\text{Erl}(x, I; s)$ is not necessarily convex.

[20] Lemma 2 will continue to hold if the expression for loss probability is derived from an $M/G/1/s$ system (Nagarajan and Towsley 1992) or is modeled as the exceedance probability of an $M/M/1$ queue.

[21] Note that these properties imply positive semidefiniteness because the Hessian is symmetric.

Now, convexity of f, together with $f(0)=0$, implies that $xf'(x)-f(x)\geq 0$. If $f(x)/I$ is convex, then $\det[\text{Hess}\, f(x)/I]=(2f(x)f''(x)-f'(x)^2)/I^4\geq 0$. The latter two expressions, together with $f(x)f'(x)\geq 0$ imply that $\det[\text{Hess}\, K(x,I)]\geq 0$. Finally, it is straightforward to check that K is strictly convex in I. $\square$

LEMMA 3. *Suppose $\ell(x,I)$ is such that $x\ell(x,I)$ is a jointly convex function of $x\geq 0$ and $I\geq 0$ on a convex domain. Let $g(y)$ be a strictly increasing, convex, and twice differentiable function of $y\geq 0$. Then $xg(\ell(x,I))$ is a jointly convex function of x and I on the same domain as ℓ. Further, if $x\ell(x,I)$ is strictly convex in x and/or I, then so is $xg(\ell(x,I))$.*

Proof. Fix $(x,I)\geq 0$ and $(x',I')\geq 0$ in the domain of ℓ, and assume at least one of x or x' is nonzero without loss of generality. Choose $\delta\in[0,1]$. Define $K(x,I)=x\ell(x,I)$, and define $y=\delta x/(\delta x+(1-\delta)x')$, and $y'=(1-\delta)x'/(\delta x+(1-\delta)x')$. Note that $0\leq y,y'\leq 1$, with $y+y'=1$. We have:

$$\begin{aligned}(\delta x+(1-\delta)x')g(\ell(\delta x+(1-\delta)x',\delta I+(1-\delta)I'))&\leq(\delta x+(1-\delta)x')g(y\ell(x,I)+y'\ell(x',I'))\\&\leq\delta xg(\ell(x,I))+(1-\delta)x'g(\ell(x',I')),\end{aligned}$$

where the first step follows by convexity of K and the fact that g is increasing, and the second follows by convexity of g and the definitions of y and y'. By the same argument, it follows that $xg(\ell(x,I))$ inherits any strict convexity properties of $x\ell(x,I)$. $\square$

We conclude with an important example that exhibits *increasing* returns to investment.

EXAMPLE 6. Suppose $\ell(x,I)$ represents the mean queueing delay of a customer in an $M/G/1$ queue in steady state, with arrival rate x, mean service time $1/I$, and service time variance σ^2/I^2.[22] In this case, the total congestion cost K exhibits increasing returns to investment. Moreover, one can show that the total cost function v is concave. As a consequence, the social optimum calls for a single firm operating.[23]

Appendix B: Proofs, Section 4

In the following lemma we show that under Assumption 4, the total cost function v is convex; note that for simplicity, in this section we omit the dependence of ℓ, K, I, and v on the subscript j. In the process, we show additional properties that will be useful in subsequent proofs. In this appendix, we use the following shorthand notation: $\partial f(x,y)/\partial x=f_x(x,y)$.

[22] Note that these conditions ensure that increasing investment expenditure reduces mean service time while holding the coefficient of variation of the service time fixed.

[23] See, for example, Smith and Whitt (1981) for results concerning queueing models where pooling resources is efficient.

LEMMA 4. *Suppose Assumption 4 holds for the congestion cost function ℓ. Then:*

1. $v(x)$ *is strictly increasing and convex for* $x > 0$ *with* $\lim_{x\to 0} v(x) = 0$*;*
2. $v'(x) = \ell(x, I(x)) + x\ell_x(x, I(x))$*; and*
3. $xv'(x) - v(x) = x^2\ell_x(x, I(x)) - I(x)$ *is nonnegative and nondecreasing for* $x > 0$*.*

Proof. That $v(x)$ is strictly increasing with $\lim_{x\to 0} v(x) = 0$ follows directly from Assumption 1. By assumption, K is convex, so $v(x)$ is convex for all $x > 0$, because it is a partial minimization of a convex function. By the envelope theorem it follows that $v'(x) = \ell(x, I(x)) + x\ell_x(x, I(x))$, where $I(x)$ is the unique efficient investment level given x.

Since v is convex with $\lim_{x\to 0} v(x) = 0$, for all $x > 0$, we have $v(x) \le xv'(x) = x\ell(x, I(x)) + x^2\ell_x(x, I(x))$. Since $v(x) = x\ell(x, I(x)) + I(x)$ by definition, we conclude that for all $x > 0$, $x^2\ell_x(x, I(x)) - I(x) \ge 0$. Finally, since v is a nonnegative convex function, $xv'(x) - v(x)$ is nondecreasing; thus $x^2\ell_x(x, I(x)) - I(x)$ is nondecreasing. □

We note in the case of constant returns to investment, the total cost function v can be shown to be *linear*, cf. Lemma 1 in Section 4.

Proof of Lemma 1. By Assumption 1, the minimization problem (5) admits an optimal solution. For $x = 0$, the optimal solution is $I = 0$. Moreover, for fixed $x > 0$, $\ell(x, I)$ is coercive in I, so any optimal solution to (5) must be interior.

Fix $x > 0$ and let I be an optimal solution. Then I must satisfy the following first order condition:

$$-\frac{x^2}{I^2}h'\left(\frac{x}{I}\right) + 1 = 0 \ . \tag{16}$$

Consider the equation:[24]

$$y^2 h'(y) = 1. \tag{17}$$

By Assumption 4, $xh(x/I)$ is strictly convex in I, hence, $y^2h'(y)$ is strictly increasing. Therefore, equation (17) has a unique solution, say ϕ. We conclude from (16) that $x/I = \phi$. This implies that for fixed x, problem (5) has a unique optimal solution, given by $I = x/\phi$; i.e., the efficient investment level is *linear* in x, and the efficient congestion level is constant over all $x > 0$, and equal to $h(\phi)$. Substituting, we conclude that $v(x) = xh(\phi) + x/\phi$, as required. □

[24] Equation (17) was also derived for a similar model studied by Xiao et al. (2007); they consider a pricing and investment game where Assumption 5 is also satisfied. However, that paper does not establish that such cost models lead to linear total cost functions.

Appendix C: Proofs, Section 5

Proof of Proposition 1. First note that the demand quantities, x^{NE}, must satisfy $x_j^{NE} > 0$ for $j \in A$. If for some $j \in A$, $x_j^{NE} = 0$, then firm j has a strictly profitable deviation by setting $I_j = 0$. Similarly, prices must satisfy $p_j^{NE} > 0$ for $j \in A$. Finally, $P(Q^{NE}) > 0$, hence, $P'(Q^{NE}) < 0$.

Without loss of generality, we assume that firm 1 is active, and consider the optimization problem for firm 1. Now, given $\{(p_k, I_k) = (p_k^{NE}, I_k^{NE})\,,\ k = 2, ..., N,\}$, it must be that

$$\begin{aligned} \left(p_1^{NE}, I_1^{NE}, \{x_k^{NE}\}_{k\in A}\right) \in \ & \underset{\substack{p_1, I_1 \geq 0 \\ x_k \geq 0,\ k \in A}}{\arg\max} \quad p_1 x_1 - I_1 \\ \text{s.t.} \qquad & p_k + \ell_k(x_k, I_k) = P(Q),\ k \in A, \end{aligned} \tag{18}$$

where $Q = \sum_{k\in A} x_k$ and we used the fact that $x_k^{NE} > 0$, for $k \in A$; and $x_k^{NE} = 0$, for $k \notin A$ (because $I_k^{NE} = 0$, for inactive firms k).

It is straightforward to check that $\left(p_1^{NE}, I_1^{NE}, \{x_k^{NE}\}_{k\in A}\right)$ satisfies the linear independence constraint qualification condition. Therefore, the Karush-Kuhn-Tucker conditions are necessary for optimality (Bertsekas et al. 2003). We know that $x_k^{NE} > 0$ for $k \in A$, $I_1^{NE} > 0$, and $p_1^{NE} > 0$. Therefore, the Lagrangian of problem (18) can be written as:

$$L(p_1, I_1, \{x_k\}_{k\in A}, \{\mu_k\}_{k\in A}) = (p_1 x_1 - I_1) - \sum_{k\in A} \mu_k \left(p_k + \ell_k(x_k, I_k) - P(Q)\right) \ .$$

The KKT conditions are given by:

$$x_1 - \mu_1 = 0\ , \quad \text{(diff. w.r.t. } p_1\text{)} \tag{19}$$

$$p_1 - \mu_1 \frac{\partial \ell_1(x_1, I_1)}{\partial x_1} + P'(Q) \sum_{k\in A} \mu_k = 0\ , \quad \text{(diff. w.r.t. } x_1\text{)} \tag{20}$$

$$1 + \mu_1 \frac{\partial \ell_1(x_1, I_1)}{\partial I_1} = 0\ , \quad \text{(diff. w.r.t. } I_1\text{)} \tag{21}$$

$$-\mu_k \frac{\partial \ell_k(x_k, I_k)}{\partial x_k} + P'(Q) \sum_{k\in A} \mu_k = 0\ ,\ k \in A,\ k \neq 1. \quad \text{(diff. w.r.t. } x_k\text{)} \tag{22}$$

Equations (19) and (22) imply that:

$$\sum_{k\in A} \mu_k = x_1 \Big/ \left(1 - \sum_{k\in A, k\neq 1} P'(Q)/(\partial \ell_k(x_k, I_k)/\partial x_k)\right),$$

which is well defined because $\partial \ell_k(x_k, I_k)/\partial x_k > 0$ for all k, $x_k > 0, I_k > 0$, and $P'(Q^{NE}) < 0$. Equation (7) follows by replacing the previous expression and equation (19) in equation (20). Equation (8) follows by replacing equation (19) in equation (21).

By definition of v_j and the Wardrop equilibrium, we see that an active firm j makes a profit $\pi(p_j^{NE}, I_j^{NE}, x_j^{NE}) = P(Q^{NE}) x_j^{NE} - v_j(x_j^{NE})$. By equation (8), we see that if Assumption 4 holds, then all active firms invest efficiently, i.e., $I_j^{NE} = I_j(x_j^{NE})$ for all $j \in A$. $\square$

Appendix D: Perfectly Inelastic Demand Model

In this section we characterize WE, social optimum, and NE when demand is perfectly inelastic and firms are homogeneous. Throughout this section we assume Assumptions 6 and 7 hold. First, we define WE.

DEFINITION 12. For given price and investment vectors $\boldsymbol{p}$ and $\boldsymbol{I}$, a vector of demand quantities $\boldsymbol{x} \geq 0$ is a *Wardrop equilibrium* if

$$p_j + \ell(x_j, I_j) = \min_k \{p_k + \ell(x_k, I_k)\} \qquad \text{for all } j \text{ with } x_j > 0; \text{and}$$
$$\sum_{j=1}^{N} x_j = M.$$

If Assumption 4 holds, then there exists a social optimum $(\boldsymbol{x}^S, \boldsymbol{I}^S)$, which is symmetric; i.e., $x_j^S = M/N$ and $I_j^S = I(M/N)$ for all j. Finally, we characterize NE in the next proposition; the proof is nearly identical to the proof of Proposition 1.

PROPOSITION 8. *Suppose Assumptions 6 and 7 hold. Suppose that the vectors of prices $\boldsymbol{p}^{NE}$, investment levels $\boldsymbol{I}^{NE}$, and demand levels $\boldsymbol{x}^{NE}$ form a 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:*

$$p_j^{NE} = x_j^{NE} \left(\frac{\partial \ell(x_j^{NE}, I_j^{NE})}{\partial x_j} + \frac{1}{\sum_{i \in A : i \neq j} \frac{1}{\partial \ell(x_i^{NE}, I_i^{NE}) / \partial x_i}} \right), \; j \in A,$$
$$0 = x_j^{NE} \frac{\partial \ell(x_j^{NE}, I_j^{NE})}{\partial I_j} + 1, \; j \in A,$$

where $Q^{NE} = \sum_{j \in A} x_j^{NE} = M$.

Appendix E: Proofs, Section 6

E.1. Uniqueness: Heterogeneous Firms and Constant Returns to Investment

We start with the following lemma, which specializes Proposition 1 to the setting of constant returns.

LEMMA 5. *Suppose Assumption 5 holds. Suppose that the vectors of prices $\boldsymbol{p}^{NE}$, investment levels $\boldsymbol{I}^{NE}$, and demand levels $\boldsymbol{x}^{NE}$ form a NE. Then the profit of firm j is $p_j^{NE} x_j^{NE} - I_j^{NE} = (P(Q^{NE}) - \xi_j) x_j^{NE}$. If firm j is active, then:*

$$P(Q^{NE}) = \xi_j + \frac{x_j^{NE}}{\sum_{i \in A : i \neq j} \phi_i x_i^{NE} - 1/P'(Q^{NE})}, \tag{23}$$

where $Q^{NE} = \sum_i x_i^{NE}$, and ϕ_j is defined as in Lemma 1. If firm j is not active, then $P(Q^{NE}) \leq \xi_j$.

Proof. The expression for firms' profits follow directly by applying Proposition 1. To show (23), note that $\partial \ell_j(x_j, I_j)/\partial x_j = h'_j(x_j/I_j)/I_j$. Since an active firm j invests efficiently at the NE, i.e., $I_j^{NE} = I_j(x_j^{NE})$, we have $\partial \ell_j(x_j^{NE}, I_j^{NE})/\partial x_j = \phi_j h'_j(\phi_j)/x_j^{NE} = 1/(\phi_j x_j^{NE})$, since $\phi_j^2 h'_j(\phi_j) = 1$ (Lemma 1). Equation (23) now follows by applying Proposition 1 and Lemma 1, and noting that $p_j^{NE} + \ell_j(x_j^{NE}, I_j^{NE}) = P(Q^{NE})$ for an active firm.

Finally, suppose that firm j is not active. and that $P(Q^{NE}) > \xi_j$, but $x_j^{NE} = I_j^{NE} = 0$. Let $\Delta^{NE} = P(Q^{NE})$ denote the full price at the Nash equilibrium. Consider a deviation where firm j attracts a small demand ϵ, and invests $I_j = I_j(\epsilon)$. In doing so, firm j is drawing demand away from other firms, and so the full price in the market will be slightly less than Δ^{NE}; we denote the resulting full price by $\Delta < \Delta^{NE}$, where $\Delta \uparrow \Delta^{NE}$ as $\epsilon \to 0$. Of course, the resulting price set by firm j must be $p_j = \Delta - \ell_j(\epsilon, I_j(\epsilon))$. Thus the profit to firm j is:

$$p_j \epsilon - I_j(\epsilon) = \Delta \epsilon - v_j(\epsilon) = (\Delta - \xi_j)\epsilon.$$

For sufficiently small $\epsilon > 0$, the latter expression will be positive; and thus firm j has a profitable deviation. We conclude that in a NE, any firm that is not active must have $\Delta^{NE} \leq \xi_j$. $\square$

Proposition 2 is now just a corollary of the preceding lemma.

Proof of Proposition 2. The result follows by the preceding lemma, since the right hand side of (23) is positive (as $P'(Q^{NE}) < 0$). $\square$

Proof of Proposition 3. Let $Q^{NE} = \sum_j x_j^{NE}$, and let $\Delta^{NE} = P(Q^{NE})$, the NE full price. Define:

$$B^{NE} = \sum_i \phi_i x_i^{NE} - \frac{1}{P'(Q^{NE})}.$$

Throughout the proof, *we restrict attention only to active firms;* denote the set of active firms by A.

Observe from Lemma 5 that for all active firms j:

$$\Delta^{NE} = \xi_j + \frac{x_j^{NE}}{B^{NE} - \phi_j x_j^{NE}}.$$

Note that $\xi_j = h_j(\phi_j) + 1/\phi_j$ (Lemma 1). Thus:

$$\begin{aligned} \Delta^{NE} &= h_j(\phi_j) + \frac{1}{\phi_j} + \frac{x_j^{NE}}{B^{NE} - \phi_j x_j^{NE}} \\ &= h_j(\phi_j) + \frac{B^{NE}/\phi_j}{B^{NE} - \phi_j x_j^{NE}}. \end{aligned}$$

If we define:

$$\Gamma_j(\Delta) = \frac{1}{\phi_j(\Delta - h_j(\phi_j))},$$

then we conclude that for any active firm j

$$\Gamma_j(\Delta^{NE}) = 1 - \frac{\phi_j x_j^{NE}}{B^{NE}}. \tag{24}$$

(Note that for all active firms j, $\Delta^{NE} > \xi_j$ from Proposition 2; in particular, this implies $\Gamma_j(\Delta^{NE}) < 1$.)

If we sum over $i \in A$, we obtain:

$$\sum_{i\in A} \Gamma_i(\Delta^{NE}) - n^* + 1 = -\frac{1}{B^{NE} P'(Q^{NE})}, \tag{25}$$

(Note that both the left and right hand sides of this equation are positive.) On the other hand, we can rearrange (24) to obtain:

$$\frac{1-\Gamma_j(\Delta^{NE})}{\phi_j} = \frac{x_j^{NE}}{B^{NE}}.$$

If we again sum over $i \in A$, we obtain:

$$\sum_{i\in A} \frac{1-\Gamma_i(\Delta^{NE})}{\phi_i} = \frac{Q^{NE}}{B^{NE}}. \tag{26}$$

We thus conclude that at any NE, the following relationship holds:

$$\frac{\sum_{i\in A} \Gamma_i(\Delta^{NE}) - n^* + 1}{\sum_{i\in A}(1-\Gamma_i(\Delta^{NE}))/\phi_i} = -\frac{1}{Q^{NE} P'(Q^{NE})}. \tag{27}$$

Now note that each function $\Gamma_i(\Delta)$ is strictly decreasing in Δ for $\Delta > \xi_i$; and the left hand side of the preceding expression is strictly increasing in each Γ_i as long as $\Gamma_i < 1$. Thus the left hand side of (27) is strictly decreasing in Δ for $\Delta > \xi_i$, for all $i \in A$.

On the other hand, we know at the NE that $P(Q^{NE}) = \Delta^{NE}$; alternatively, in terms of the demand function, $D(\Delta^{NE}) = Q^{NE}$. Further, by definition, $P'(Q) = 1/D'(P(Q))$ for all Q, so we conclude that the right hand side of the expression (27) is equal to $-D'(\Delta^{NE})/D(\Delta^{NE})$. Further, we know that $D(\Delta^{NE}) > 0$. Thus if D is log-concave over the region where it is positive, then $-D'(\Delta^{NE})/D(\Delta^{NE})$ a nondecreasing function of Δ^{NE}. Thus there exists at most one solution Δ^{NE} to (27). This in turn ensures that Q^{NE} is uniquely determined, and by (25) or (26), we conclude that B^{NE} is uniquely determined as well. Finally, (24) shows that x_j^{NE} is uniquely determined as well, and thus I_j^{NE} is uniquely determined (since $I_j^{NE} = I_j(x_j^{NE})$). This fixes all active firms' prices as well, since $p_j^{NE} = \Delta^{NE} - \ell_j(x_j^{NE}, I_j^{NE})$. $\square$

Proof of Proposition 4. Suppose that $(\boldsymbol{p}^{NE}, \boldsymbol{x}^{NE}, \boldsymbol{I}^{NE})$ and $(\overline{\boldsymbol{p}}^{NE}, \overline{\boldsymbol{x}}^{NE}, \overline{\boldsymbol{I}}^{NE})$ are two distinct equilibria, and let A and $\overline{A}$ denote the set of active firms respectively. Let $Q^{NE} = \sum_j x_j^{NE}$, and $\Delta^{NE} = P(Q^{NE})$; define $\overline{Q}^{NE}$ and $\overline{\Delta}^{NE}$ similarly. Since both are threshold equilibria, assume without loss of generality that $\overline{A}$ is a strict subset of A. (In other words, $\overline{A} = \{1, \ldots, \overline{n}^*\}$, and $A = \{1, \ldots, n^*\}$, with $n^* > \overline{n}^*$.) We derive a contradiction.

We use the same notation as the proof of Proposition 3. Recall from (24) that at a NE, $\Gamma_i(\Delta^{NE}) < 1$ for all active firms i. Thus:

$$\sum_{i \in \overline{A}} \Gamma_i(\Delta^{NE}) - \overline{n}^* + 1 > \sum_{i \in A} \Gamma_i(\Delta^{NE}) - n^* + 1,$$

and:

$$\sum_{i \in \overline{A}} (1 - \Gamma_i(\Delta^{NE}))/\phi_i < \sum_{i \in A} (1 - \Gamma_i(\Delta^{NE}))/\phi_i.$$

We rewrite (27) using the demand function D, as follows:

$$\frac{\sum_{i \in A} \Gamma_i(\Delta^{NE}) - n^* + 1}{\sum_{i \in A} (1 - \Gamma_i(\Delta^{NE}))/\phi_i} = -\frac{D'(\Delta^{NE})}{D(\Delta^{NE})}. \tag{28}$$

We conclude that:

$$\frac{\sum_{i \in \overline{A}} \Gamma_i(\Delta^{NE}) - \overline{n}^* + 1}{\sum_{i \in \overline{A}} (1 - \Gamma_i(\Delta^{NE}))/\phi_i} > -\frac{D'(\Delta^{NE})}{D(\Delta^{NE})}.$$

As discussed in the proof of Proposition 3, note that the left hand side of the preceding expression is strictly decreasing in Δ^{NE}, and the right hand side is nondecreasing in Δ^{NE}, over the region of interest. Since an analog of (28) must hold for $\overline{\Delta}^{NE}$ and $\overline{n}^*$ as well, we conclude that $\overline{\Delta}^{NE} > \Delta^{NE}$.

But now recall from Proposition 2 that a firm i is active in a NE if and only if $\Delta^{NE} > \xi_i$. Thus if $\overline{\Delta}^{NE} > \Delta^{NE}$, then any firm that is in A must also be in $\overline{A}$—a contradiction. We conclude that the set of active firms in any NE is uniquely determined, as required.

Finally, Assumption 3 implies $P(0) > \xi_1$, hence, firm 1 must be active in any equilibrium. □

Proof of Theorem 1. The result follows directly from Propositions 2, 3, and 4. We just need to show that $Q^{NE} < Q^S$. Note that the socially optimal outcome allocates all demand to firm 1, and sets $P(Q^S) = \xi_1$ (since firm 1 has the lowest total cost coefficient). On the other hand, by Proposition 2, we must have $P(Q^{NE}) > \xi_1$. since P is strictly decreasing where positive, we conclude $Q^{NE} < Q^S$. □

E.2. Uniqueness: Homogeneous Firms and Nonincreasing Returns to Investment

In this section, we use the following shorthand notation: $\partial f(x,y)/\partial x = f_x(x,y)$.

Proof of Proposition 5. We show that it must be the case that $x_i = x_j$, for all $i, j \in A$, where A is the set of active firms. By Proposition 1, this suffices to establish that a NE is symmetric among active firms. Fix $i, j \in A$ with $i \neq j$.

Recalling the WE and NE conditions (equations (2) and (7)), it is useful to define:

$$g(x_i,x_j) \equiv \ell(x_i,I(x_i)) + x_i\ell_x(x_i,I(x_i)) + \frac{x_i}{B+1/\ell_x(x_j,I(x_j))} \\ - \left(\ell(x_j,I(x_j)) + x_j\ell_x(x_j,I(x_j)) + \frac{x_j}{B+1/\ell_x(x_i,I(x_i))}\right) ,$$

where $B=\sum_{k\in A,k\neq i,j} 1/\ell_x(x_k,I(x_k)) - 1/P'\left(\sum_{k\in A} x_k\right) > 0$. Since $\boldsymbol{x}$ are the NE demand quantities, it must be that $g(x_i,x_j)=0$.

Suppose that $x_i > x_j$ to arrive to a contradiction. We do this in two steps using two different monotonicity arguments over $g(x_i,x_j)$. For the first step, recall that $v(x)$ is convex (Lemma 4), hence $v'(x)=\ell(x,I(x))+x\ell_x(x,I(x))$ is nondecreasing. Therefore, if $x_i>x_j$ and $g(x_i,x_j)=0$, it must be that $\ell_x(x_i,I(x_i)) > \ell_x(x_j,I(x_j))$.

For the second step, we regroup terms and rewrite $g(x_i,x_j)$ as follows:

$$g(x_i,x_j) = \ell(x_i,I(x_i)) - \ell(x_j,I(x_j)) + \frac{K_1K_2}{K_3K_4},$$

where

$$K_1 = x_i - x_j + B\left(x_i\ell_x(x_i,I(x_i)) - x_j\ell_x(x_j,I(x_j))\right);$$

$$K_2 = \ell_x(x_i,I(x_i)) + \ell_x(x_j,I(x_j)) + B\ell_x(x_i,I(x_i))\ell_x(x_j,I(x_j));$$

$$K_3 = B\ell_x(x_i,I(x_i)) + 1;$$

$$K_4 = B\ell_x(x_j,I(x_j)) + 1.$$

Since $x_i > x_j > 0$ and $\ell_x(x_i,I(x_i)) > \ell_x(x_j,I(x_j)) > 0$, we have $x_i\ell_x(x_i,I(x_i)) > x_j\ell_x(x_j,I(x_j))$. Also, by assumption $\ell(x_i,I(x_i)) \geq \ell(x_j,I(x_j))$. This implies that $g(x_i,x_j) > 0$, a contradiction. Thus we must have $x_i = x_j$, as required. $\square$

LEMMA 6. *Suppose Assumption 4 holds. If for all $x>0$ and $I>0$,* (11) *holds:*

$$\frac{\partial}{\partial I}\mathrm{MRS}(x;I) \geq \frac{1}{x},$$

then the effective congestion cost function $\ell(x,I(x))$ is nondecreasing for all $x>0$.

Proof. We first define the following notation. For $x>0$, let $J(x;L)$ denote the L-isoquant of ℓ; i.e., $J(x;L)$ is the unique solution to: $\ell(x,J(x;L)) = L$. Note that by Assumption 1, if $L_1 < L_2$, then $J(x;L_1) > J(x;L_2)$. Further, the implicit function theorem establishes that: $J'(x;L) = \mathrm{MRS}(J(x;L);x)$. In other words, the slope of the L-isoquant is exactly the marginal rate of substitution. So now suppose $L_1 < L_2$.

By integrating both sides of (11) from $I = J(x;L_2)$ to $I = J(x;L_1)$, we conclude that if $L_1 < L_2$, then:

$$J'(x;L_1) - J'(x;L_2) \geq \frac{1}{x}\left(J(x;L_1) - J(x;L_2)\right). \qquad (*)$$

(Note this conclusion only holds when $J(x;L_1) > J(x;L_2)$.)

Define $W(x,L) = xL + J(x,L)$. Since $K(x,I)$ is strictly convex in I, there exists a unique investment level $I(x)$ for x. Let $L_2 = \ell(x, I(x))$, and choose $L_1 < L_2$. Note that $J(x;L_2) = I(x)$. For fixed $x > 0$, we have:

$$\begin{aligned}
\frac{\partial W}{\partial x}(x,L_1) - \frac{\partial W}{\partial x}(x,L_2) &= L_1 + J'(x;L_1) - L_2 - J'(x;L_2) \\
&\geq \frac{1}{x}\left(xL_1 + J(x;L_1) - xL_2 + J(x;L_2)\right) \\
&= \frac{1}{x}\left(W(x,L_1) - W(x,L_2)\right) > 0,
\end{aligned}$$

where the first inequality follows by $(*)$, and the second inequality follows since $I(x)$ is the unique efficient investment given x. Thus for all sufficiently small $\epsilon > 0$, we conclude that for all $L_1 < L_2$:

$$W(x+\epsilon, L_1) - W(x+\epsilon, L_2) > 0. \qquad (**)$$

Let $l(x+\epsilon, I(x+\epsilon)) = L_\epsilon$. Since $I(x+\epsilon)$ is the efficient investment level, we must have $W(x+\epsilon, L_\epsilon) \leq W(x+\epsilon, L_2)$. Combined with the inequality $(**)$, we conclude $L_\epsilon \geq L_2$, i.e., that $l(x+\epsilon, I(x+\epsilon)) \geq l(x, I(x))$ for all sufficiently small $\epsilon > 0$. This concludes the proof. $\square$

Proof of Proposition 6. Our proof involves two steps. We first show active firms' profits are strictly positive in a symmetric NE. We then use this fact to show an inactive firm has a profitable deviation by entering.

We first show active firms make positive profits in a symmetric NE. Consider the monopoly setting where a firm's profits can be shown to be equal to $P(x)x - v(x)$ when demand x is served. The monopolist makes positive profits because for $x > 0$ small enough, $P(x) > v'(x) \geq v(x)/x$. The first inequality follows by Assumptions 2 and 3. The second inequality follows by Lemma 4, because $v(x)$ is convex with $\lim_{x\to 0} v(x) = 0$. Thus we assume at least two firms are active.

By Lemma 4, $x^2 \ell_x(x, I(x)) - I(x) \geq 0$ for all $x > 0$. In addition, note that at a symmetric NE with $n > 1$ active firms, by (9) we have $p_j^{NE} > x_j^{NE} \ell_x(x_j^{NE}, I(x_j^{NE}))$ for all active firms j. Combining the last two inequalities implies that active firms make positive profits in a symmetric NE.

Finally, suppose a symmetric NE only has $n < N$ firms active. Let P_n^{NE} the price set by each active firm and Q_n^{NE} the total mass of consumers served. Consider an inactive firm that considers attracting a share ϵ of the market and denote $Q(\epsilon)$ the corresponding total mass of consumers served. Note that by continuity, $\lim_{\epsilon\to 0} Q(\epsilon) = Q_n^{NE}$. In this case the profits of the firm will be:

$$A(\epsilon) = \left[P_n^{NE} + \ell\left(\frac{Q(\epsilon)-\epsilon}{n}, I\left(\frac{Q_n^{NE}}{n}\right)\right) - \ell(\epsilon, I(\epsilon))\right]\epsilon - I(\epsilon). \tag{29}$$

The first term on the right represents the full price after this inactive firm enters, less the congestion cost experienced by any consumers of this firm; this is the maximum price the new entrant can charge. The last term is the investment expenditure. We have:

$$\liminf_{\epsilon\to 0}\frac{A(\epsilon)}{\epsilon} = P_n^{NE} + \ell\left(\frac{Q_n^{NE}}{n}, I\left(\frac{Q_n^{NE}}{n}\right)\right) - \liminf_{\epsilon\to 0}\frac{v(\epsilon)}{\epsilon}.$$

Since v is convex with $v(\epsilon)\to 0$ as $\epsilon\to 0$ (by Lemma 4), the last term in the preceding expression is at most $v(Q_n^{NE}/n)/(Q_n^{NE}/n)$. Thus:

$$\liminf_{\epsilon\to 0}\frac{A(\epsilon)}{\epsilon} \ge P_n^{NE} + \ell\left(\frac{Q_n^{NE}}{n}, I\left(\frac{Q_n^{NE}}{n}\right)\right) - \frac{v(Q_n^{NE}/n)}{Q_n^{NE}/n} > 0,$$

where the final inequality follows because all active firms make positive profit at the NE, and $v(x) = x\ell(x, I(x)) + I(x)$. Therefore, a new entrant has a profitable deviation for sufficiently small $\epsilon > 0$, concluding the proof. □

Proof of Theorem 2. We prove that $Q^{NE} < Q^S$; the remainder of the theorem follows by Proposition 5, Proposition 6, and Lemma 6.

Let x^S and x^{NE} be socially optimal and NE demand allocations, respectively. The social planner's first order conditions are given by: $P(Q^S) = v'(x_i^S)$. Replacing p_j^{NE} and I_j^{NE} from equations (7) and (8) in (2), we find: $P(Q^{NE}) > \ell\left(x_j^{NE}, I\left(x_j^{NE}\right)\right) + x_j^{NE}\ell_x\left(x_j^{NE}, I\left(x_j^{NE}\right)\right) = v'(x_i^{NE})$, where the inequality follows since ℓ is strictly increasing in x and $P'(Q^{NE}) < 0$ (Proposition 1). By assumption P is a strictly decreasing function, while by Lemma 4 v' is an increasing function. We conclude that $Q^{NE} < Q^S$, as desired. □

Corollary 1. *Suppose the congestion cost function takes one of the forms in Lemma 2. Then the conclusions of Theorem 2 hold.*

Proof. We prove the lemma in the case where $\kappa(I) = I$; the generalization for κ concave is straightforward, and we omit the details. We need to check that the congestion cost functions in Lemma 2 satisfy the hypothesis of Theorem 2. That Assumptions 1 and 4 hold follow from Lemma 2. It remains to be shown that (11) holds.

Case 1: $\ell(x, I) = f(x/I)$. It is simple to check that, in this case, $\partial\,\mathrm{MRS}(x; I)/\partial I = 1/x$.

Case 2: $\ell(x, I) = g\left(\mathrm{Erl}(x, I; s)\right)$. Condition (11) is satisfied by the same argument as in case 1.

Case 3: $\ell(x, I) = f(x)/I$, *where* $f(x)/I$ *is a convex function.* In this case, $\partial\,\mathrm{MRS}(x; I)/\partial I = f'(x)/f(x)$. Since f is a convex function with $f(0) = 0$, $f'(x)/f(x) \ge 1/x$. Hence, condition (11) is satisfied. □

Appendix F: Proofs, Section 7

F.1. Existence: Heterogenous Firms

Proof of Theorem 3. We prove existence of NE by applying Kakutani's fixed point theorem. Let $BR(\boldsymbol{p}_{-1}, \boldsymbol{I}_{-1})$ be the set of best response price and investment levels of firm 1, given the prices and investment levels of every other firm. That is, $BR(\boldsymbol{p}_{-1}, \boldsymbol{I}_{-1})$ is the set of solutions to the following optimization problem, given $(\boldsymbol{p}_{-1}, \boldsymbol{I}_{-1})$:

$$\begin{aligned}
\text{maximize} \quad & p_1 x_1 - I_1 \\
\text{subject to} \quad & \Delta \leq p_j + \ell_j(x_j, I_j), \text{ for all } j; \\
& \Delta = p_j + \ell_j(x_j, I_j), \text{ for all } j \text{ s.t. } x_j > 0; \\
& P(Q) = \Delta; \\
& \sum_j x_j = Q; \\
& p_1, I_1, \boldsymbol{x}, \Delta, Q \geq 0.
\end{aligned}$$

Because $P(q)$ is concave and strictly decreasing, it must be that $\lim_{q\to 0} P(q) < \infty$ and $P(\hat{q}) = 0$, for large enough $\hat{q}$. Therefore, a best response will always be an element of a compact strategy space, so we can restrict attention to that set. In addition, by Berge's maximum theorem the correspondence BR is non-empty and upper-hemicontinuous.

To finish the proof, we show that $BR(\boldsymbol{p}_{-1}, \boldsymbol{I}_{-1})$ is a convex set, for every $(\boldsymbol{p}_{-1}, \boldsymbol{I}_{-1})$. For some values of $(\boldsymbol{p}_{-1}, \boldsymbol{I}_{-1})$, $BR(\boldsymbol{p}_{-1}, \boldsymbol{I}_{-1}) = \{(p, 0) : p \geq 0\}$, so it is convex. Let us assume that $(\boldsymbol{p}_{-1}, \boldsymbol{I}_{-1})$ is such that all elements of $BR(\boldsymbol{p}_{-1}, \boldsymbol{I}_{-1})$ are of the form (p, I), with $I > 0$. In this case, firm 1's optimization problem is equivalent to:

$$\begin{aligned}
\text{maximize} \quad & \Delta x_1 - v_1(x_1) \\
\text{subject to} \quad & \Delta \leq p_j + \ell_j(x_j, I_j), \text{ for all } j; \\
& \Delta = p_j + \ell_j(x_j, I_j), \text{ for all } j \text{ s.t. } x_j > 0; \qquad (30) \\
& \sum_j x_j = D(\Delta); \\
& \boldsymbol{x}, \Delta \geq 0.
\end{aligned}$$

Let $x_j(\Delta)$ be the unique solution to equation (30) if $\Delta > p_j$, and define $x_j(\Delta) = 0$ for $\Delta \leq p_j$. Since $\ell_j(x, I)$ is concave and strictly increasing is x for each $I > 0$, $x_j(\Delta)$ must be convex and nondecreasing in Δ.

When the full price is Δ, there holds $D(\Delta) = x_1 + \sum_{j>1} x_j(\Delta)$; let $x_1(\Delta)$ denote the solution x_1 to this equation:

$$x_1(\Delta) = D(\Delta) - \sum_{j>1} x_j(\Delta).$$

Note that $D(\Delta)$ is only strictly positive on an interval $\Delta \in [0, \Delta^{\max})$. Further, firm 1 only optimizes over $x_1 > 0$. For this reason, we restrict attention to the set of full prices $S = \{\Delta < \Delta^{\max} : x_1(\Delta) > 0\}$. Over $\Delta \in S$, the function $x_1(\Delta)$ is concave and strictly decreasing. We conclude that $x_1(\Delta)$ has a concave and strictly decreasing inverse $\delta(x_1)$, defined over the domain $(0, D(0))$. (Note that $D(0)$ is the maximum possible mass of consumers that could be served by firm 1.)

Thus for $x_1 \in (0, D(0))$, we can write the profit of firm 1 in terms of the demand x_1 as follows:

$$\pi(x_1) = x_1 \delta(x_1) - v_1(x_1).$$

Since δ is concave and strictly decreasing on $(0, D(0))$, it follows that the profit of firm 1 is concave as a function of x_1. Hence, the set of maximizers X_1^* of $\pi(x_1)$ is convex. Moreover, $(p_1, I_1) \in BR(\boldsymbol{p}_{-1}, I_{-1})$, if and only if $p_1 = \delta(x_1) - \ell_1(x_1, I_1(x_1)) \equiv p_1(x_1)$ and $I_1 = I_1(x_1)$, for some $x_1 \in X_1^*$. The mappings p_1 and I_1 are continuous. Hence $BR(\boldsymbol{p}_{-1}, I_{-1})$ is a convex set. The result follows by Kakutani's fixed point theorem. □

Proof of Proposition 7. Our proof follows in two stages. First, relying on the proof of Proposition 3, we show that there exists a candidate NE that satisfies the local optimality conditions for each firm identified in Lemma 5. Second, we show that no firm has a profitable unilateral deviation from this candidate equilibrium, as required.

By assumption, $\lim_{\Delta \to \overline{\xi}} A(\Delta) > \lim_{\Delta \to \overline{\xi}} B(\Delta)$ and $\overline{\xi} < \overline{\Delta}$; in particular, this ensures $D(\overline{\xi}) > 0$. Since $D(\Delta)$ is concave and strictly decreasing where positive, $\overline{\Delta}$ must be finite. Further, $B(\Delta) \to \infty$ as $\Delta \to \overline{\Delta}$, but $A(\Delta)$ remains bounded as $\Delta \to \overline{\Delta}$. Following the proof of Proposition 3, note that in the region where $\Delta > \overline{\xi}$ and where $D(\Delta)$ is positive, $A(\Delta)$ is strictly decreasing in Δ, and $B(\Delta)$ is nondecreasing in Δ. By continuity, there must exist a solution Δ^* such that $A(\Delta^*) = B(\Delta^*)$, and $\overline{\xi} < \Delta^* < \overline{\Delta}$. Note that since $\xi_i = h_i(\phi_i) + 1/\phi_i$, and $\Delta^* > \xi_i$, we have $0 < \Gamma_i(\Delta^*) < 1$.

Let $Q^* = D(\Delta^*) > 0$. Define B^* using the analog of either (25) or (26):

$$\sum_i \Gamma_i(\Delta^*) - N + 1 = -\frac{1}{B^* P'(Q^*)}; \quad \sum_i \frac{1 - \Gamma_i(\Delta^*)}{\phi_i} = \frac{Q^*}{B^*}.$$

Either equation yields the same value B^*, since $A(\Delta^*) = B(\Delta^*) = -1/(Q^* P'(Q^*))$. Finally, for each firm i, define x_i^* using the analog of (24):

$$x_i^* = \frac{B^*(1 - \Gamma_i(\Delta^*))}{\phi_i}.$$

(Note that the procedure we used to derive x_i^* ensures that $\sum_i x_i^* = Q^*$.) For each firm i, define the price $p_i^* = \Delta^* - h_i(\phi_i)$; and define the investment level $I_i^* = I_i(x_i^*) = x_i^*/\phi_i$.

We now prove that the prices $\boldsymbol{p}^*$, investment levels $\boldsymbol{I}^*$, and demand allocations $\boldsymbol{x}^*$ constitute a NE with all firms active. Following Proposition 1 and Lemma 5, it is straightforward to verify that (by construction) these vectors are *locally stationary* for each firm: that is, the change in profit to any firm for an infinitesimal deviation away from the candidate equilibrium is zero; we omit the details. To complete the proof, therefore, it suffices to show that in fact, no firm has *any* profitable deviation from the candidate equilibrium.

First, observe that in this candidate equilibrium, each firm makes positive profits, since the profit of firm i is $(\Delta^* - \xi_i)x_i^* > 0$. Thus, no firm would deviate to make itself *inactive*. We restrict attention, therefore, to considering deviations by firm i where firm i remains active. Further, as before, an active firm always invests efficiently given the demand it serves; thus even if firm i deviates and serves demand x_i instead, its profit will be $(\Delta - \xi_i)x_i$. Finally, observe that since firm i remains active in any deviation, we can assume without loss of generality that $\Delta < P(0)$ in any deviation. We thus have the following optimization problem for firm i:

$$\begin{aligned}
\text{maximize} \quad & (\Delta - \xi_i)x_i && (31)\\
\text{subject to} \quad & \Delta \le p_j^* + h_j(x_j/I_j^*), \text{ for all } j \ne i; && (32)\\
& \Delta = p_j^* + h_j(x_j/I_j^*), \text{ for all } j \ne i \text{ s.t. } x_j > 0; && (33)\\
& \sum_j x_j = D(\Delta); && (34)\\
& x_i > 0, \boldsymbol{x}_{-i} \ge 0, \xi_i < \Delta < P(0). && (35)
\end{aligned}$$

Since $h_j(x_j/I_j^*)$ is strictly increasing in x_j, and $(\Delta - \xi_i)x_i$ is strictly increasing in x_i where $\Delta > \xi_i$, we can eliminate (33); in an optimal solution, if $x_j > 0$, then (32) must bind. In addition, we can replace the equality in (34) by an inequality. In an optimal solution the inequality must bind; if not the objective function can grow by increasing x_i. Taking the logarithm of the objective function yields:

$$\begin{aligned}
\text{maximize} \quad & \log(\Delta - \xi_i) + \log x_i\\
\text{subject to} \quad & \Delta \le p_j^* + h_j(x_j/I_j^*), \text{ for all } j \ne i;\\
& \sum_j x_j \le D(\Delta);\\
& x_i > 0, \boldsymbol{x}_{-i} \ge 0, \xi_i < \Delta < P(0).
\end{aligned}$$

We now introduce a new variable $z_i = \log(\Delta - \xi_i)$, to obtain:

$$\begin{aligned} \text{maximize} \quad & z_i + \log x_i && (36) \\ \text{subject to} \quad & z_i \le \log(p_j^* - \xi_i + h_j(x_j/I_j^*)), \text{ for all } j \neq i; && (37) \\ & \sum_j x_j \le D(\exp(z_i) + \xi_i); \\ & x_i > 0, \boldsymbol{x}_{-i} \ge 0, z_i > -\infty. && (38) \end{aligned}$$

Observe that if D is concave and strictly decreasing, then $D(\exp(z_i)+\xi_i)$ is concave as well. Further, note that if $K + h(y)$ is log-concave for some positive constant K, then for any $k < K$ and $I > 0$, $k + h(y/I)$ is log-concave over the region where it is positive. Thus, given our assumptions on h_j and the fact that $p_j^* < \overline{\Delta}$, it can be verified that the right hand side of the constraint (37) is a concave function of x_j. As a result, the preceding optimization problem has a concave objective function and a convex feasible region, so any locally stationary solution must be globally optimal.

To conclude the proof, it suffices to observe that the fact that the candidate NE is locally stationary for firm i's original optimization problem (including the binding WE constraint (33)) implies it must be locally stationary for the problem (36)-(38) as well, as any local profitable deviation in the latter problem yields a local profitable deviation in the original problem. We conclude that no firm has a profitable deviation from the candidate Nash equilibrium, as required. □

F.2. Existence: Homogenous Firms

Proof of Theorem 4. We employ a technique identical to the proof of Proposition 7. Since firms are homogeneous, we suppress any dependence of h, ξ, ϕ, or Γ on the firm index i. First, we show that there exists a candidate NE full price $\Delta^* > \xi$ such that $A(\Delta^*) = B(\Delta^*)$, where $A(\cdot)$ and $B(\cdot)$ are defined as in Proposition 7. We then show that in fact no firm has a profitable deviation from this candidate NE, as required. Throughout this proof, we adopt the same notation as the proof of Proposition 7.

We can simplify $A(\Delta)$ as follows:

$$A(\Delta) = \phi\left(\frac{1}{N(1-\Gamma(\Delta))} - 1\right). \qquad (39)$$

Observe that as $\Delta \downarrow \xi$, we have $\Gamma(\Delta) \uparrow 1$, and thus $A(\Delta) \to \infty$. On the other hand, since $D(\xi) > 0$ (cf. Assumption 3), $B(\Delta)$ remains bounded as $\Delta \downarrow \xi$. We can thus conclude by a similar argument to the proof of Proposition 7 that there exists a unique solution Δ^* with $\xi < \Delta^* < \overline{\Delta}$, such that $A(\Delta^*) = B(\Delta^*)$. As in that proof, we obtain a candidate NE $\boldsymbol{p}^*, \boldsymbol{I}^*, \boldsymbol{x}^*$ with full price Δ^* and total demand Q^* that is locally stationary for every firm i. Note that (by construction) this NE will be symmetric, and that all firms will be active.

The second step is to ensure that in fact, no firm has any profitable deviation from the candidate equilibrium. The proof is identical to the proof of Proposition 7; the only difference is that now, we show directly that the right hand side of (37) is concave at the candidate NE. To do this, we first compute $\Delta^* - \xi$. Observe that since $\xi = 1/\phi + h(\phi)$,

$$\frac{1}{1-\Gamma(\Delta^*)} = 1 + \frac{1}{\phi(\Delta^* - \xi)}.$$

Thus, using (39) and the fact that $A(\Delta^*) = B(\Delta^*)$, we conclude that:

$$\Delta^* - \xi = \frac{1}{(N-1)\phi - ND'(\Delta^*)/D(\Delta^*)} < \frac{1}{(N-1)\phi - ND'(\xi)/D(\xi)},$$

where the inequality follows because $\Delta^* > \xi$ and D is concave where it is positive.[25] Thus, given equation (12) together with the fact that $\phi^2 h'(\phi) = 1$, it follows after some algebra that $\Delta^* - \xi < h(\phi)$. Since $\Delta^* = p_j^* + h(\phi)$ at the candidate equilibrium, we conclude that $p_j^* - \xi < 0$. To conclude the proof, observe that if $h(y)$ is log-concave, then for $k < 0$, $k + h(y/I)$ is log-concave where it is positive. This ensures that the right hand side of (37) is concave, and the remainder of the proof follows as in Proposition 7, as required. $\square$

In the remainder of the appendix, we use the following shorthand notation: $\partial f(x,y)/\partial x = f_x(x,y)$.

Proof of Theorem 5. It suffices to establish the candidate NE specified in Theorem 2 is in fact a NE. Without loss of generality, assume that firms $j = 2, \ldots, N$ choose the same investment level $I_j = I^{NE}$, and price $p_j = P^{NE}$ given by the NE in Theorem 2. We already know that in this case, $p_1 = P^{NE}$ and $I_1 = I^{NE}$ satisfy the first order conditions for profit maximization for firm 1; we must show that in fact these choices are optimal for firm 1. From Proposition 6, firm 1 would make positive profits in this case; so we can assume without loss of generality that firm 1 optimizes only over price and investment pairs that yield positive demand x_1.

Although in this case it is not generally true that $\pi(x_1)$ is concave (like in Theorem 3), we can show that $\pi(x_1)$ is *quasiconcave*; i.e., π is monotonically nondecreasing for $x \le M/N$, and monotonically nonincreasing for $x \ge M/N$. This will imply that the candidate NE is in fact a NE.

For notational simplicity, define $g(y) = \ell(y, I(M/N))$, and $h(y) = yg(y)$. The proof rests on investigating the properties of $g(y) - (M/(N-1) - y)g'(y)$.

[25] Note that the equality could also be derived directly by using (23) and the fact that all firms are homogeneous and active, and the candidate NE is symmetric.

Step 1: $g(y)/g'(y) - (M/(N-1) - y)$ is a nondecreasing function of $y \geq 0$. This follows immediately from our assumption that $\ell_x(x,I)/\ell(x,I)$ is nonincreasing in x.

Step 2: $g(M/N)/g'(M/N) - (M/(N-1) - M/N) \geq 0$. This inequality is equivalent to $(M/N)g'(M/N)/g(M/N) + 1 \leq N$, which is again true by assumption.

Step 3: $\pi(x_1)$ is concave and nondecreasing for $x_1 \leq M/N$. Since $\pi'(M/N) = 0$, it suffices to show concavity for $x_1 \leq M/N$. In this case, we can write firm 1's profit as:

$$\pi(x_1) = \left(P^{NE} + \ell\left(\frac{M-x_1}{N-1}, I\left(\frac{M}{N}\right)\right)\right)x_1 - v(x_1), \tag{40}$$

Differentiate (40):

$$\pi'(x_1) = P^{NE} + g\left(\frac{M-x_1}{N-1}\right) - \left(\frac{x_1}{N-1}\right)g'\left(\frac{M-x_1}{N-1}\right) - v'(x_1). \tag{41}$$

Since v is convex (cf. Lemma 4), $v'(x)$ is nondecreasing. Thus if we substitute $y = (M-x_1)/(N-1)$, it suffices to show that $g(y) - (M/(N-1) - y)g'(y)$ is nondecreasing in y for $y \geq M/N$ (since $x_1 \leq M/N$). Since g is convex, $g'(y)$ is nondecreasing; thus by combining Steps 1 and 2, we conclude $g(y) - (M/(N-1) - y)g'(y)$ is nondecreasing.

Step 4: $\pi'(x_1) \leq 0$ for $x_1 \geq M/N$. Substituting the definition of P^{NE} from (10) into (41), we have:

$$\pi'(x_1) = \left(\frac{M}{N-1}\right)\left(g'\left(\frac{M}{N}\right) - g'\left(\frac{M-x_1}{N-1}\right)\right) + h'\left(\frac{M-x_1}{N-1}\right) - v'(x_1).$$

By convexity, $v'(x_1) \geq v'(M/N)$ if $x_1 \geq M/N$. Further, by Lemma 4, we have $v'(M/N) = \ell(M/N, I(M/N)) + (M/N)\ell_x(M/N, I(M/N)) = h'(M/N)$. Thus defining $y = (M-x_1)/(N-1)$, we have for $x_1 \geq M/N$:

$$\pi'(x_1) \leq \left(\frac{M}{N-1}\right)\left(g'\left(\frac{M}{N}\right) - g'(y)\right) - \left(h'\left(\frac{M}{N}\right) - h'(y)\right). \tag{$**$}$$

Note that if $x_1 \geq M/N$, we have $y \leq M/N$. From Steps 1 and 2, together with the fact that $g'(y)$ is nondecreasing and positive, we conclude that for $y \leq M/N$:

$$g(y) - \left(\frac{M}{N-1} - y\right)g'(y) \leq g\left(\frac{M}{N}\right) - \left(\frac{M}{N-1} - \frac{M}{N}\right)g'\left(\frac{M}{N}\right).$$

Rearranging terms and recalling that $h(y) = yg(y)$, the preceding inequality establishes that the right hand side of ($**$) is nonpositive, as required.

The preceding two steps ensure that $\pi(x_1)$ is quasiconcave. We conclude firm 1 maximizes profit when $x_1 = M/N$, which is ensured exactly through the choice $p_1 = P^{NE}$ and $I_1 = I^{NE} = I(M/N)$, as required. □

Appendix G: Proofs, Section 8

Proof of Theorem 6. For the purposes of this theorem, we assume the number of firms N can take values over the entire half line $(0,\infty)$. We show that conditions (1), (2) , and (3) from Mankiw and Whinston (1986) stated in the text hold in our model.

1. Q_N is strictly increasing in N and $\lim_{N\to\infty} Q_N = \overline{Q} < \infty$. By equations (2) and (9), and Lemma 4, it must be that in the unique and symmetric NE:

$$\begin{aligned} 0 &= P(Q_N) - v'(q_N) - \frac{q_N}{(N-1)/\ell_x(q_N, I(q_N)) - 1/P'(Q_N)} \\ &= P(Q_N) - \xi - \frac{P'(Q_N)Q_N}{\phi(N-1)P'(Q_N)Q_N - N} \\ &\equiv H(Q_N, N), \end{aligned} \tag{42}$$

where the second equation follows by Lemma 1. By the implicit function theorem:

$$\frac{\partial Q_N}{\partial N} = -\frac{\partial H(Q,N)/\partial N}{\partial H(Q,N)/\partial Q}. \tag{43}$$

It is straightforward to check that if the inverse demand function is concave, then $\partial H(Q,N)/\partial N > 0$ and $\partial H(Q,N)/\partial Q < 0$. Hence, Q_N is strictly increasing in N. Finally, by Theorem 2 we know that $Q_N < Q^S$, for all N. Therefore, $P'(Q_N)Q_N < \infty$, for all N, hence, $\lim_{N\to\infty} P(Q_N) - \xi = 0$. It follows that $\lim_{N\to\infty} Q_N = Q^S < \infty$.

2. q_N is strictly decreasing in N. We know that $q_N = Q_N/N$, hence, $\partial q_n/\partial N = (N\partial Q_N/\partial N - Q_N)/(N^2)$. Note that $\lim_{N\to 0} Q_N > 0$ (see equation (42)). By differentiating (43) it is straightforward to check that $\partial^2 Q_N/\partial N^2 < 0$, hence, Q_N is strictly concave. It follows that $N\partial Q_N/\partial N - Q_N < 0$, so q_N is strictly decreasing in N.

3. $P(Q_N) - v'(q_N) \geq 0,\ \forall N$. This follows directly from equations (2) and (9).

Finally, by Assumption 9, $\Pi(1) \geq 0$. In addition, $\lim_{N\to\infty} P(Q_N) - \xi = 0$, hence $\lim_{N\to\infty} \Pi(N) = -F < 0$. Moreover,

$$\Pi'(N) = q_N P'(Q_N)\partial Q_N/\partial N + P(Q_N)\partial q_N/\partial N - v'(q_N) < 0.$$

Hence, the free entry equilibrium number of firms exist and is unique. □

Appendix H: Entry: Inelastic Demand

In this section, we assume that demand is perfectly inelastic of size M. *Hence, throughout this section we suppose Assumption 7 holds.* The analysis is similar to Section 8, but there are some notable differences. As previously mentioned, the analysis in Mankiw and Whinston (1986) cannot be applied in this context, so our approach here is different to Section 8. In addition, we more generally assume nonincreasing returns to investment. Finally, since demand is inelastic of size M in our asymptotic analysis it is more natural to increase the market size M to infinity, instead of decreasing the sunk entry cost to zero.

As a benchmark, we first consider the problem faced by a social planner that chooses the number of participant firms in the industry. The social objective is to maximize total social surplus. Since demand is perfectly inelastic and the unique NE is symmetric and efficient, this is equivalent to minimizing the sum of congestion costs, investment expenditures, and entry costs for a symmetric demand allocation with efficient investment levels.[26]

Definition 13. A number of firms N^S is socially optimal if it maximizes total social surplus for a symmetric demand allocation with efficient investment levels, i.e. if it solves:

$$\text{minimize} \quad S(N) = M\ell\left(\frac{M}{N}, I\left(\frac{M}{N}\right)\right) + NI\left(\frac{M}{N}\right) + NF \ . \tag{44}$$

Recall that in this case $\Pi(N) = \pi(N) - F$, where $\pi(N)$ is the post-entry profit to a firm in the symmetric and efficient NE when there are N incumbent firms; see Theorem 2 for $N \geq 2$. Monopoly profits, $\Pi(1)$, are assumed to be infinite.

First, we compare the free entry equilibrium with the socially optimal number of firms and show that, in general, there is excessive free entry. To derive the result we need an additional assumption.

Assumption 10. *The total congestion cost $K(x,I)$ is submodular in x and I, for $x, I > 0$, i.e. $K_{xI}(x,I) \leq 0$ for $x, I > 0$.*

The assumption is satisfied by all the congestion models studied in Lemma 2. We have the following theorem.

Theorem 8. *Suppose Assumptions 4, 6, 7, 8, and 10 hold. Then the free entry equilibrium number of firms exists and is unique, and it is no smaller than one less any socially optimal number of firms.*

Proof. First, we prove the following lemma which will be useful to prove this and the next theorem. For the purposes of this lemma, we treat $S(N)$ and $\Pi(N)$ as functions defined over the entire half line $[2, \infty)$.

[26] According to the literature's terminology, the *first-best number of firms* maximizes total social surplus assuming that the social planner can also decide the behavior of firms in the post-entry game. Recall that the *second-best number of firms* maximizes total social surplus assuming that the social planner can only decide the number of firms, but cannot affect post-entry behavior of firms. Note that because the second stage post-entry NE is efficient (Theorem 2), then these two solution concepts coincide.

Lemma 7. *Suppose Assumptions 4, 6, and 7 hold. Then, $S(N)$ is convex, and $\lim_{N\to\infty}\Pi(N)<0$. If in addition Assumption 10 holds, then the profit function $\Pi(N)$ is nonincreasing for $N\geq 2$.*

Proof of Lemma. We prove that $S(N)$ is convex. Note that $S(N)=Nv(M/N)+NF$, where v is defined as in (5). Thus:

$$S'(N)=v(M/N)-(M/N)v'(M/N)+F, \tag{45}$$

which is nondecreasing by Lemma 4. Thus S is convex.

We now prove that $\lim_{N\to\infty}\Pi(N)<0$. Let P_N^{NE} denote the symmetric NE price with $N\geq 2$ incumbent firms; i.e., P_N^{NE} is defined as in (10). It suffices to show that the NE price is uniformly bounded (i.e., $\sup_{N\geq 2}P_N^{NE}<\infty$), since in that case $\lim_{N\to\infty}\Pi(N)\leq -F$. To see this, note from Lemma 4 that $v'(x)$ is nondecreasing, so $\sup_{x\leq M/2}v'(x)<\infty$. Substituting for $v'(x)$, this implies $\sup_{x\leq M/2}x\ell_x(x,I(x))<\infty$, which implies in turn that $\sup_{N\geq 2}P_N^{NE}<\infty$.

Now suppose that Assumption 10 holds; we prove that $\Pi(N)$ is nonincreasing for $N\geq 2$. Using (45) and (10), we conclude that for $N\geq 2$:

$$\Pi(N)=\frac{N}{N-1}(F-S'(N))+\frac{I\left(M/N\right)}{N-1}-F. \tag{+}$$

By (45) and Lemma 4, $F-S'(N)$ is nonincreasing and nonnegative, so the first term is nonincreasing for $N\geq 2$. Using standard supermodularity results (Topkis 1998), Assumption 10 implies that the efficient investment level $I(x)$ is a nondecreasing function. Therefore, the second term in the right hand side of equation (+) is nonincreasing. We conclude that $\Pi(N)$ is nonincreasing for $N\geq 2$. □

If $N^S=1$ or $N^S=2$, $N^E\geq N^S-1$ follows directly, because $N^E\geq 1$. Therefore, we assume without loss of generality that $N^S>2$.

First, we prove that a free entry equilibrium number of firms exists and is unique. By assumption monopoly profits are arbitrarily large, hence, $\Pi(1)>0$. By Lemma 7, $\Pi(N)$ is nonincreasing for $N\geq 2$, and $\lim_{N\to\infty}\Pi(N)<0$. Hence, a free entry equilibrium exists and is unique.

Now, we show that the free entry equilibrium number of firms is no smaller than one less any socially optimal number of firms. We first show that $\Pi(N)+S'(N)>0$. To see this note that for $N\geq 2$:

$$S'(N)=-\Pi(N)+\left(\frac{M}{N^2}\right)P_N^{NE}, \tag{46}$$

where P_N^{NE} is the NE price with N firms defined as in (10). This result follows from (45), Lemma 4, and equation (10). Since $P_N^{NE}>0$, we conclude $\Pi(N)+S'(N)>0$.

Let N^E be the free entry equilibrium number of firms and N^S the largest socially optimal number of firms. We want to prove that $N^E \geq N^S - 1$. To provide intuition, we first ignore the integrality constraints on N^E and N^S. In this case, it must be that $S'(N^S) = 0$ and $\Pi(N^E) = 0$. From the preceding paragraph, it follows that $\Pi(N^S) > 0$. Therefore, since $\Pi(N)$ is nonincreasing, $N^E \geq N^S$.

We now refine the argument to cope with the integrality constraint. Since $S(N)$ is convex (Lemma 7), there must exist $N \geq 2$ such that $S'(N) = 0$, and either $N^S = \lceil N \rceil$ or $N^S = \lceil N \rceil - 1$. (Here $\lceil x \rceil$ denotes the ceiling of x.) As before, we can conclude $\Pi(N) > 0$, so $N^E \geq N^S - 1$, since $\Pi(N)$ is nonincreasing. □

We note that Scotchmer (1985) proves the same result for club goods assuming that the sunk entry cost, F, equals zero. Scotchmer's result is vacuous in our setting, because if $F = 0$, the free entry equilibrium and the socially optimal number of firms are not well defined.

Next, we show that if a condition satisfied by many models of interest holds, entry becomes asymptotically efficient as the market size grows. We consider a sequence of industries indexed by the market size, M. The congestion cost function, ℓ, and the sunk entry cost, F, remain the same for all market sizes. We extend the previous definitions of total social cost $S(N)$ and profits $\Pi(N)$ to allow for dependence on the market size. Hence, we let $S(N, M)$ and $\Pi(N, M)$ denote the total social cost and symmetric NE profits, respectively, when there are N firms and the market size is equal to M. Let $N^S(M)$ and $N^E(M)$ be a socially optimal and the free entry equilibrium number of firms, respectively, for market size M.

We first note that without any additional conditions, as the market size grows the free entry equilibrium number of firms may be significantly larger than the socially optimal number of firms. For example, suppose $\ell(x, I) = x/I$. In this case it is simple to check that the unique socially optimal number of firms is $N^S(M) = 1$, for all M, and that the free entry equilibrium number of firms is given by $N^E(M) = \frac{1}{2}\left(1 + \sqrt{1 + 4M/F}\right)$. As a consequence, $\lim_{M\to\infty} N^E(M)/N^S(M) = \infty$.[27]

We will show an asymptotic efficiency result under the assumption that the total congestion cost, $K(x, I) = x\ell(x, I)$, is jointly strictly convex in x and I. In addition, we will assume the fixed cost is not excessively large; if it were, then the social optimum will involve only one firm operating regardless of the market size, but the free entry equilibrium number of firms may be significantly larger. Formally, we define $G(x) = xv'(x) - v(x)$, where v is defined as in (5). It is simple to show

[27] Note that also $\lim_{F\to 0} N^E(F) = \infty$ and $N^S(F) = 1$, for all $F > 0$. Recall that in Section 8 where demand is downward sloping we obtained $\lim_{F\to 0} N^E(F) = \infty$ and $\lim_{F\to 0} N^S(F) = \infty$ (Theorem 7). The difference is that even under constant returns to investment, in the second best problem with a downward sloping demand function, the socially optimal number of firms increases as F becomes small, so that the total number of consumers served increases, increasing social surplus.

that if $K(x,I)$ is jointly strictly convex in x and I and $F \geq \sup_x G(x)$, then $S(N,M)$ is strictly increasing in $N \geq 1$ for all M, which implies $N^S(M) = 1$ for all M. To ensure entry is asymptotically large, we thus also assume $\sup_x G(x) > F$.

Theorem 9. *Suppose Assumptions 4, 6, 7, 8, and 10 hold. In addition, suppose the total congestion cost $K(x,I)$ is jointly strictly convex in x and I. Suppose also that $\sup_x G(x) > F$, where $G(x) = xv'(x) - v(x)$.*

Then the socially optimal number of firms $N^S(M)$ and the free entry equilibrium number of firms $N^E(M)$ grow in proportion to the market size asymptotically as the market size grows. As a consequence, $\sup_M N^E(M)/N^S(M) < \infty$. Further, entry becomes asymptotically efficient:

$$\lim_{M\to\infty} \frac{S(N^E(M),M) - S(N^S(M),M)}{S(N^S(M),M)} = 0. \tag{47}$$

Proof. To simplify the analysis we will ignore the integer constraints on the number of entrants. Since we are considering asymptotic results where the number of entrants grow to infinity, this will not have any effect on our results. Thus we must have $S_N(N^S(M),M) = 0$ (assuming $N^S(M)$ is interior) and $\Pi(N^E(M),M) = 0$. The proof follows several steps. First, we prove that for all sufficiently large M, $N^S(M)$ is linear in M. Then, we prove that $N^E(M)$ grows proportional to M as $M \to \infty$. Finally, we show that entry becomes asymptotically efficient.

By equation (45), $S_N(N^S(M),M) = F - G(M/N)$. Since $\sup_x G(x) > F$, for M large enough, $S_N(1,M) < 0$, hence, $N^S(M)$ must be interior. In this case, $S_N(N^S(M),M) = 0$, and it follows that:

$$G\left(\frac{M}{N^S(M)}\right) = F.$$

From Lemma 4 it follows that G is nondecreasing. It is straightforward to show that when K is strictly convex, then v is strictly convex as well; and this in turn suffices to ensure that G is in fact strictly increasing when K is strictly convex.

Since, in addition, $\sup_x G(x) > F$ with $G(x) \to 0$ as $x \to 0$, there must exist a unique ξ^* such that $G(\xi^*) = F$. Thus we conclude that for all M large enough we have $N^S(M) = \xi^* M$, i.e., $N^S(M)$ is linear in M for all sufficiently large M.

We now consider asymptotic behavior of $N^E(M)$. By Theorem 8, $N^E(M)$ cannot grow slower than M as $M \to \infty$. On the other hand, if $N^E(M)$ grows faster than M as $M \to \infty$, then $\lim_{M\to\infty} \Pi(N^E(M),M) < 0$ by an argument similar to the proof that $\lim_{N\to\infty} \Pi(N) < 0$ in Lemma 7. Since $\Pi(N^E(M),M) = 0$, we conclude $N^E(M)$ grows in proportion to M as $M \to \infty$.

Finally, we prove that entry becomes asymptotically efficient. By convexity of $S(N,M)$ in N (cf. Lemma 7), we know:

$$S(N^E(M),M) - S(N^S(M),M) \leq S_N(N^E(M),M)\left(N^E(M) - N^S(M)\right), \tag{48}$$

Since $\Pi(N^E(M),M) = 0$, by equation (46) we have:

$$S_N(N^E(M),M) = \frac{M}{(N^E(M))^2} P^{NE}(N^E(M),M),$$

where $P^{NE}(N,M)$ is the NE price defined as in (10). Since $N^E(M)$ grows in proportion to M as $M \to \infty$, it follows from (10) that $S_N(N^E(M),M)$ decreases in proportion to $1/M$ as $M \to \infty$. We conclude the right hand side of (48) remains bounded as $M \to \infty$. The limit (47) follows, because $S(N^S(M),M)$ grows in proportion to M asymptotically as $M \to \infty$. $\square$

We note that the conditions that $K(x,I)$ is jointly strictly convex and $\sup_x G(x) > F$ are satisfied, for example, by any congestion function of the form $\ell(x,I) = x^p/I^q$ with $p > 1,\ p > q$. In this case, $G(x) \to \infty$ as $x \to \infty$. Note that if $p = q = 1$, $G(x) = 0$ and K is convex but not strictly convex.

The result implies that as the market size grows, the free entry equilibrium number of firms grows to infinity, the industry becomes fragmented, and market shares converge to zero. Also, the free entry equilibrium number of firms becomes asymptotically efficient. A heuristic argument suggests the latter result. Recall that if the symmetric NE price was equal to the Pigovian price, then the free entry equilibrium number of firms would be socially optimal. As the market size grows, the free entry equilibrium number of firms grows proportional to the market size. Thus the difference between the symmetric NE price and the congestion externality converges to zero, ensuring firms' entry decisions are asymptotically efficient.

Acknowledgments

We 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, Informs, Kellogg, MIT, and Stanford. We 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 grant numbers 0428868 and IIS-0428868, and by the Okawa Foundation.

References

Acemoglu, D., K. Bimpikis, A. Ozdaglar. 2008. Price and capacity competition. To appear in *Games and Economic Behavior.*

Acemoglu, D., A. Ozdaglar. 2007. Competition and efficiency in congested markets. *Mathematics of Operations Research* **32**(1) 1–31.

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

Allon, G., A. Federgruen. 2008. Service competition with general queueing faciltites. *Operations Research* **56**(4) 827–849.

Baake, P., K. Mitusch. 2007. Competition with congestible networks. *Journal of Economics* **91**(2) 151–176.

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

Bertsekas, D. P., A. Nedić, A. E. Ozdaglar. 2003. *Convex Analysis and Optimization*. 1st ed. Athena Scientific.

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

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

De Borger, B., K. Van Dender. 2006. Prices, capacities, and service levels in a congestible bertrand duopoly. *Journal of Urban Economics* **60** 264–283.

Deneckere, R., J. Peck. 1995. Competition over price and service rate when demand is stochastic: a strategic analysis. *The Rand Journal of Economics* **26**(1) 148–162.

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

Engel, E. M., R. Fischer, A. Galetovic. 2004. Toll competition among congested roads. *Topics in Economic Analysis & Policy* **4**(1).

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

Harel, A. 1990. Convexity properties of the Erlang loss formula. *Operations Research* **38**(3) 499–505.

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

Johari, R., G. Y. Weintraub. 2008. Competition and contracting in service industries. Working paper, Stanford University.

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

Kreps, D. M., J. A. Scheinkman. 1983. Quantity precommitment and Bertrand competition yield cournot outcomes. *The Bell Journal of Economics* **14**(2) 326–337.

Levitan, R., M. Shubik. 1972. Price duopoly and capacity constraints. *International Economic Review* **13**(1) 111–122.

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

Mankiw, N. G., M. D. Whinston. 1986. Free entry and social inefficiency. *The Rand Journal of Economics* **17**(1) 48–58.

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

Nagarajan, R., D. Towsley. 1992. A note on the convexity of the probability of a full buffer in the M/M/1/K queue. CMPSCI Technical Report TR 92-85, University of Massachusetts at Amherst.

Ozdaglar, A. 2008. Price competition with elastic traffic. *Networks* .

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

Roughgarden, Tim. 2005. *Selfish Routing and the Price of Anarchy*. MIT Press, Cambridge, Massachusetts.

Scotchmer, S. 1985. Profit-maximizing clubs. *Journal of Public Economics* **27** 25–45.

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

Smith, D. R., W. Whitt. 1981. Resource sharing for efficiency in traffic systems. *The Bell System Technical Journal* **60**(1) 39–55.

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

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

Topkis, D. 1998. *Supermodularity and Complementarity*. 1st ed. Princeton University Press.

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

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

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

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