Competition in wireless telecommunications

Miguel Angel Campo-Rembado and Arun Sundararajan[1]

Leonard N. Stern School of Business, New York University
44 West 4th Street, New York, NY 10012
(mcampo, asundara) @stern.nyu.edu

Abstract: This paper presents an analysis of competition in wireless telecommunications that models the interdependence between spectrum availability, network infrastructure deployment, the generation of transmission technology, average traffic levels and service quality. We show that the constraints on spectrum availability and infrastructure that characterize this industry lead to service quality levels that are endogenously affected by market share, and this negative externality leads to bilaterally higher pricing power for competing providers. We incorporate the effects of these negative usage externalities into a two-stage game of quality competition with required minimum infrastructure levels, and establish two distinct kinds of fulfilled-expectations subgame perfect equilibria. Under the first equilibrium, which occurs at both very low and very high levels of average traffic, providers deploy the minimum permissible network infrastructure and price symmetrically. The second kind of equilibrium is asymmetric in both network deployment levels and pricing, though the equilibrium extent of quality differentiation is moderated by the externalities highlighted earlier. Analysis of these equilibria reveals three phases in a wireless market's evolution. In early-stage markets, providers should maintain minimum infrastructure levels and avoid active quality-based differentiation, relying instead on externality-based pricing power. As wireless markets mature, providers need to pursue an aggressive quality differentiation strategy, accompanied by continuous and rapid growth of their network infrastructure. Our model explains why the observed industry trend of relatively flat average revenue per user may be a natural equilibrium outcome during this phase, even though usage levels and value grow steadily. Finally, we identify a threshold level of average per-user traffic at which viable service quality levels and positive profits are not sustainable, and discuss how the rapidly declining profits and quality levels that precede this threshold should trigger active migration to the next generation of transmission technology.

[1] This draft: August, 2004. We thank Lee Sproull, Michael Ward, Kevin Zhu, and seminar participants at New York University, the 15th Workshop on Information Systems and Economics, and the London Business School's 'Competition in Networking' conference. This research has been supported in part by the La Caixa Fellowship Foundation.

1. Introduction

This paper develops a model of competition which captures characteristics of supply and demand that are specific to wireless telecommunications, towards prescribing pricing, network deployment and migration strategy based on a lucid economic model of the underlying technology, and also towards explaining observed industry trends more accurately. Our analysis is likely to apply to future telecommunications and networking industries that share similar characteristics. In the process, we extend the standard model of quality-differentiated competition to incorporates negative usage externalities and minimum capacity constraints, an extension which may be pertinent to addressing many related research problems in IT economics.

Service quality in wireless telecommunications is an important basis for consumer choice. Surveys of the wireless industry, however, suggest that there is not much variation in the quality levels actually chosen by different providers, and that pricing tends to be fairly uniform; currently, each major U.S. provider has a menu of pricing plans which start with a minimum price of about thirty dollars, and measured price per minute-of-usage shows only small variations between the major U.S. providers. In contrast, standard models of quality-differentiated competition (beginning with Shaked and Sutton, 1982) posit that since an increase in similarity in product quality increases the threat of price competition, equilibria always feature substantially different prices and quality levels. Additionally, equilibrium prices are predicted to increase when there is a uniform increase in product value across customers. This conflicts with observed behavior in the wireless industry, where despite an increase in the range of features as well as the average usage of consumers, total prices for wireless service (as measured by ARPU: average revenue per user per month, a common industry measure) have not changed significantly[2].

Some technology-specific features of supply and demand in this industry may explain these deviations. First, wireless service quality depends on a set of inter-related technological choices, rather than being a simple, directly chosen strategic variable. Firms can influence quality by varying their

[2] For example, according to CTIA's semi-annual cellular survey, average minutes of usage per user grew over 20% from 2002 to 2003, while average revenue per user was almost flat, increasing by less than 3%, which is roughly in pace with the CPI inflation rate.

deployed level of network infrastructure; this is a 'short-run' variable to some extent, and firms are continually investing in additions and upgrades to their networks. Additionally, at a fixed level of network infrastructure, quality is influenced by a pair of related technological choices – the amount of spectrum the firm owns, and the effectiveness of utilization of this spectrum by the type (generation) of technology used by the firm. These tend to be 'long-run' choices: a shift to a new generation of transmission technology is a multi-billion dollar undertaking, requiring an overhaul of the firm's network infrastructure and simultaneous upgrades in consumer hardware; firms are also often restricted by regulatory constraints on spectrum availability and trading.

A second distinguishing feature of the wireless industry is the presence of specific kind of negative usage externality. An increase in network traffic increases the fraction of traffic to an transmission tower that cannot be carried (and is therefore 'lost'), and service quality, measured as a function of this loss rate, is therefore endogenously affected by the equilibrium market shares of competing providers. For a fixed set of technology choices, the magnitude of these externalities are also influenced by the average traffic generated by each user, which has risen steadily over time, driven by increased voice calling and the adoption of features like web access, email, gaming and text messaging.

Furthermore, each firm's *minimum* choice of quality is indirectly restricted by a number of factors. A minimum level of network infrastructure is needed to cover a region at any non-zero level of quality, since cell phone battery limitations, interference and certain kinds of signal distortion limit the maximum area that a single transmission tower can cover. In some cases, firms granted licenses in the 'beauty contests' for wireless spectrum are required by regulators to pre-commit to minimum levels of population coverage and network deployment. Additionally, most wireless service providers carry a substantial portfolio of debt, and the associated bond covenants may require them to own a minimum level of network assets. Therefore, while quality itself may not directly have a lower bound, there is likely to be a minimum level of network infrastructure required from any viable provider.

In our model, quality of the wireless service offered by each firm is therefore influenced by these three technological choices – network infrastructure, spectrum availability and effectiveness of spectrum utilization (the generation of transmission technology). Each competing firm is required to deploy a minimum level of network infrastructure. The service displays negative usage externalities stemming from increased loss rates as demand increases. Equilibrium strategies are influenced by an exogenous average level of traffic generated by each user in the market. We relate each of these factors to a single-dimensional measure of quality, based on a simplified model of how wireless transmission actually works. For an exogenously specific generation of technology and a fixed amount of spectrum, we base our results on a game in which providers make sequential network infrastructure choices and pricing decisions, and market share expectations of consumers are fulfilled in equilibrium.

Our analysis begins by showing that the presence of negative usage externalities lowers the intensity of price competition in the wireless industry. For any fixed difference in quality levels, these externalities increase the slopes of each firm's profit function, leading to a bilateral increase in pricing power. A similar effect has been observed in exogenous-capacity models of competing service systems subject to congestion (Levhari and Luski, 1978), and in models of delivery-time-based competition (Lederer and Li, 1997), though their externalities stem from waiting times in a queue, rather than a loss of traffic. We decompose equilibrium prices into a *quality premium* and an *externality premium*, and characterize some pricing and revenue paths in terms of this interpretation.

Next, we establish that our two-stage game of network deployment and pricing has two kinds of subgame perfect equilibria that satisfy fulfilled expectations. Under the first kind, both firms deploy the minimum levels of network infrastructure, leading to prices and quality levels that are symmetric. As one might expect, this equilibrium occurs at very low levels of average traffic. Additionally, this is also the equilibrium at very *high* levels of average traffic. Intuitively, this is because when network traffic is sufficiently high, the negative usage externalities reduce the marginal quality impact of additional network infrastructure to a point where it is too low to justify its cost.

In contrast, under the second kind of equilibrium, which occurs over a wide range of intermediate levels of average traffic, providers deploy asymmetric network infrastructure levels. Service quality is always vertically differentiated across firms, and equilibrium prices are asymmetric. This is similar to the equilibrium one would expect from a model of quality competition, though the presence of externalities moderates the level of differentiation to some extent. Prices stay relatively flat over a wide range of average traffic levels; this is consistent with the industry trend we highlighted earlier. Moreover, the profits of the firm with higher service quality are sometimes lower than those of the lower-quality firm.

Our analysis of these equilibria suggests three phases in a wireless market's evolution, each with its own prescribed technology deployment and pricing strategy. In early-stage markets, when the average traffic levels generated by each customer's usage are low, competing providers should deploy the minimum required network infrastructure and keep their quality levels similar, since the externality premium provides each with a significant amount of pricing power. As traffic per user increases and the market matures, the externality premium drops rapidly; therefore, aggressive quality differentiation and infrastructure growth is necessary to sustain pricing power. Finally, at a threshold level of average traffic, the impending reversion to the symmetric minimum-infrastructure equilibrium suggests a natural transition point beyond which migrating to a new generation of technology is necessary to remain viable. These strategies are described in further detail in Sections 4 and 5.

The literature on competition in the wireless industry is not extensive; most related work in this area has focused on competition in wireline telecommunications (for instance, the articles in Spulber, 1995), auction design or spectrum bidding strategies (for instance, Ausubel et al., 1997), and interconnection pricing (Laffont, Rey and Tirole, 1998, Armstrong, 1998, Mendelson and Shneorson, 2003). The quality-based nature of competition in wireless telecommunications has been recognized by prior papers, including Reiffen et. al. (2000), who base their empirical study of the regulatory implications of vertical integration in the wireless industry on a standard model of vertical differentiation. Our results contrast with those of Valetti's (1999) model of capacity pricing, in which coverage, rather than loss rates, serve as a proxy for quality. His analysis is based on the assertion that all carriers have substantial excess capacity; however, a more informed assessment of the technology underlying wireless transmission suggests that network capacity and spectrum bandwidth are important constraints and drivers of quality. Besides, coverage levels for competing wireless providers tend to be comparable in most major markets.

Our results also highlight the influence of negative usage externalities in determining wireless service quality and equilibrium pricing/infrastructure decisions. Negative externalities are widely recognized as being characteristic of information systems, and critical to IT pricing and capacity decisions. A number of papers have studied their effects, though mostly in monopoly markets, in a variety of IT-based contexts. These include models of negative externalities in shared computing resources (Mendelson, 1985, Mendelson and Whang, 1990), service facilities (Dewan and Mendelson, 1990), information systems services (Westland, 1992), electronic data interchange (Wang and Seidmann, 1995), real-time databases (Konana, Gupta and Whinston, 2000), IT resources within firms (Nadiminti, Mukhopadhyay and Kreibel, 2002), business-to-business intermediaries (Bhargava and Choudhary, 2004), peer-to-peer networks (Asvanund, Clay, Krishnan and Smith, 2004) and Internet caching services (Hosanagar, Krishnan, Chuang and Choudhary, 2003). We add to this literature; additionally, our model of quality-differentiated duopoly with externalities may facilitate future studies of competition in many of the other contexts mentioned above.

Related papers that model negative externalities in a duopoly setting with exogenously specified capacity include Luski (1976), Levhari and Luski (1978), Lederer and Li (1997), and Armony and Haviv (2003). The latter models do not consider endogenous infrastructure (capacity) investments; clearly, since their context is different from ours, they also do not explicitly model vertical differentiation or capture those aspects of supply and demand in the wireless market that we have chosen to highlight. A model by Reitman (1991), who examines simultaneous choice of price and capacity for competing firms with congestion effects is somewhat similar to ours, though his focus is mainly on the case of perfect competition. Moreover, our minimum network deployment constraint induces a second kind of symmetric duopoly equilibrium that does not arise in unconstrained models of quality differentiation.

The rest of this paper is organized as follows. Section 2 outlines the model and provides a preliminary analytical description of the effects of usage externalities on quality and payoffs. Section 3 describes the fulfilled-expectations subgame perfect equilibrium of the deployment-pricing game, establishing some important properties of both symmetric and asymmetric equilibria. Section 4 analyzes the variation in pricing, network deployment, profits and quality, and discusses the associated prescribed strategy in three market phases, each of which corresponds to an increasing range of traffic per user. Section 5 summarizes the paper's managerial implications and outlines directions for future research.

2. Model

2.1. Firms and customers

There are two firms A and B (henceforth called *providers*) who provide wireless communication services (henceforth called the *service*) over a pre-specified geographic region, whose area is normalized to 1. The customers in this area each generate a homogeneous level of *average traffic* from their usage of the service, which is denoted E, and is measured in Erlangs (a standard measure of the 'amount' of demand in communications services[3]). The physical locations of the customers are assumed to be uniformly distributed over the area. The total number of customers per unit area is normalized to 1.

Each provider is assumed to have a combination of spectrum bandwidth and transmission technology that results in their having a total of v_i *effective channels*[4]. Provider i makes two choices – their level of network infrastructure, which we model as the number of *base stations* N_i they deploy per unit area, and the total price p_i they charge each customer for their service. Each base station is assumed to have a single transmitter that can utilize all of the v_i effective channels. The number of base stations per unit area directly determines the number of *cells* that carrier i divides the area into. Since customers are uniformly distributed in the region, we also assume that the deployment of the base stations is uniform across the area, and that all cells are of identical area[5]. With N_i base

[3]A population generates a traffic of z Erlangs when, on average, the population will demand the resources (in this case, *effective channels*, to be defined shortly) required to conduct an average of z calls per unit time.

[4]For instance, if a carrier has total spectrum of 3MHz (3000 kHz), and transmission technology is TDMA, which requires 30kHz of spectrum per channel, and allows three simultaneous send-recieve transmissions (using time-division multiplexing) on each 30kHz block, this results in a total of $\frac{3000\times 3}{30} = 300$ effective channels. On the other hand, if GSM is the transmission technology, it requires the same 30kHz blocks, but allows upto 8 simultaneous send-receive pairs per block (due to superior time-division multiplexing), this results in $\frac{3000\times 8}{30} = 800$ effective channels. For the CDMA protocol, the computation of effecitive channels is based on the Shannon formula.

[5]For simplicity, we ignore corner issues, and any issues of efficient network topology. By assuming uniform base station deployment and therefore symmetric coverage, we are precluding the possibility of horizontal differentiation that might arise out of heterogeneity in the firms' relative coverage of different regions.

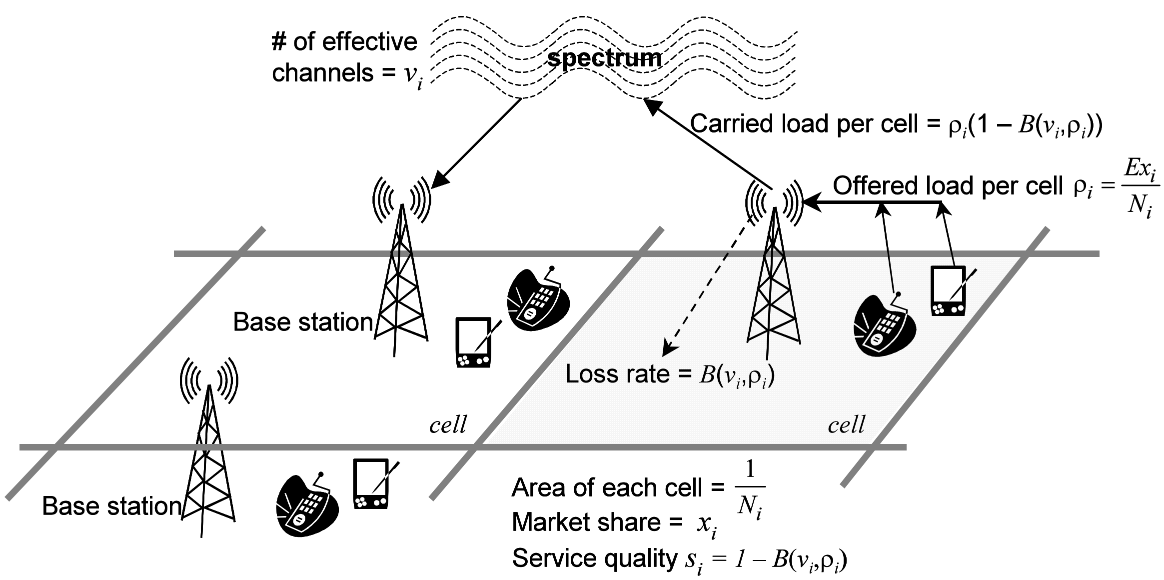


Figure 2.1: Illustration of some of the model's variables, for provider i

stations per unit area, the area of each cell is therefore $\frac{1}{N_i}$.

As a consequence, given N_i and E, if provider i has market share x_i, then the offered load of demand on the base station in each cell is

$$\rho_i = \frac{Ex_i}{N_i}.$$

Service quality is assumed to be proportional to the fraction of demand that is successfully carried, and is determined by the *loss rate* at each base station[6]. Specifically, since the base station in each cell has access to v_i channels, and has a total offered load ρ_i, the service quality for provider i takes the form:

$$s_i = 1 - B(v_i, \rho_i),$$

where $B(\rho, v)$ is the Erlang loss function, defined as:

$$B(v, \rho) = \frac{\rho^v / v!}{\sum_{i=1}^{v} (\rho^i / i!)}.$$

[6]The primary driver of quality for voice communication is the fraction of dropped calls, which is directly related to the loss rate. Other aspects to quality include interference, noise, voice distortion and periodic silences, all of which are also related (though not entirely) to loss rates. In wireless data networks, the primary measure of quality is transmission speed, which is again directly related to the fraction of lost packets (since each lost packet must be resent, thereby reducing throughput and the effective network speed).

The Erlang loss function $B(\rho, v)$ is the loss rate from an $M/G/v/v$ queue with offered load ρ. It is widely used to model loss rates in telephony and data networks[7] (Jagerman, Melamed and Willinger, 1996).

Customers differ in their valuation of service quality s. Specifically, at a level of service quality s, and at a total price p, the preferences of a customer of type θ are represented by the utility function $U(s, \theta, p)$:

$$U(s, \theta, p) = u(E)[w + \theta s] - p. \tag{2.1}$$

w represents the common (quality independent) value that every customer derives from the service (for instance, the value of having a cellular telephone in case of emergencies[8]). θ is assumed to have an absolutely continuous distribution $F(\theta)$ with support $[0, 1]$, and is assumed to be uniformly distributed in Section 3. The function $u(E)$ is strictly positive, non-decreasing and strictly concave in E, and models the increase in value to customers as the homogeneous average level of demand E increases. We assume that w is high enough for the market to be saturated in equilibrium, and all customers buy from one or the other provider.

The fixed costs to the provider are assumed to be primarily related to deploying network infrastructure. Specifically, both providers have identical cost functions $c(N)$, where N is the number of base stations deployed. These costs per base station include networking equipment, civil work, permission acquisition, and fixed line costs to connect the base station to the land-based backbone network. For simplicity, we ignore interconnection pricing issues with the land-based network and the competing network. $c(N)$ is assumed to be strictly increasing and (weakly) convex in N. In our discussion of Section 4, $c(N)$ should be interpreted as the periodic *expense* associated with the depreciation of the investment in network infrastructure of size N.

We also assume that each firm is required to deploy at least at a minimum level $N_{\min}$ of network infrastructure. As discussed briefly, there are many reasons for this lower bound on network deploy-

[7] An implicit assumption from choosing the Erlang loss function is that the region is sufficiently densely populated to approximate the demand process as a Poisson arrival process.

[8] Since the computation of E is customarily made at peak load, the term w may also capture the value that all customers get from a fraction of their off-peak usage, if this fraction is sufficiently low.

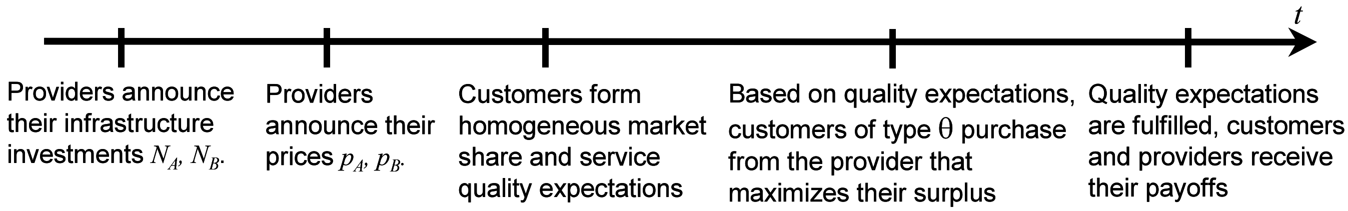


Figure 2.2: Timeline of events

ment. From a technological point of view, firms need to deploy a minimum number of base stations to cover any given area, since cell phone battery limitations, interference and Doppler distortion limit the maximum cell size, which in turn imposes a constraint on the minimum number of base stations required to offer service (of non-zero quality) to a given area. From a financial standpoint, wireless service providers are typically highly leveraged, and their bond covenants often require them to have deployed a minimum level of network assets. There are also minimum coverage and deployment requirements that the providers who are granted licenses for wireless spectrum through 'beauty contests' need to comply with.

2.2. Customer expectations and service quality

The timeline of events is summarized in Figure 2.2. In the analysis that follows, we assume that the two providers have identical spectrum bandwidth and transmission technology, and therefore, $v_A = v_B = v$. Suppose the providers have announced their capacity N_A and N_B, and their prices p_A and p_B. Without loss of generality, since the providers are identical, let $p_A \geq p_B$. Define $y \in [0, 1]$ as the expected indifferent customer type. The variable y is also referred to as the *market-share expectation* of the customers, since given y, the expected market share of provider A is $[1 - F(y)]$, and the expected market share of provider B is $F(y)$. Therefore, the expected service quality of the providers is:

$$s_A = 1 - B(v, \frac{E(1 - F(y))}{N_A}) \tag{2.2}$$

$$s_B = 1 - B(v, \frac{EF(y)}{N_B}) \tag{2.3}$$

If $p_A > p_B$, the indifferent customer type must be one for whom $U(s_A, y, p_A) = U(s_B, y, p_B)$, which simplifies to:

$$y = \frac{p_A - p_B}{u(E)(s_A - s_B)} \tag{2.4}$$

Equations (2.2), (2.3) and (2.4) define the fulfilled-expectations equilibrium demand and quality. Any solution (y^*, s_A^*, s_B^*) to these three simultaneous equations is a fulfilled-expectations equilibrium consistent with the choices p_A, p_B, N_A and N_B.

2.3. The effects of usage externalities on pricing power

Before characterizing the game and its equilibria in Section 3, we qualitatively analyze the effects of the usage externality on equilibrium pricing behavior. Fix the capacity investments N_A and N_B, and denote the expected quality difference between the providers as

$$\Delta s = s_A - s_B = B(v, \frac{EF(y)}{N_B}) - B(v, \frac{E(1 - F(y))}{N_A}). \tag{2.5}$$

Equations (2.4) and (2.5) implicitly define a quality difference correspondence $\Delta s(p_A, p_B)$. Assuming that this is single-valued, the second-period payoff functions for the two providers are:

$$\pi^A(p_A, p_B) = p_A[1 - F(y)], \tag{2.6}$$
$$\pi^B(p_A, p_B) = p_B F(y), \tag{2.7}$$

where

$$y = \frac{p_A - p_B}{u(E)\Delta s(p_A, p_B)}. \tag{2.8}$$

(2.6), (2.7) and (2.8) yield the following equations:

$$\pi_1^A(p_A, p_B) = [1 - F(y)] - \left[p_A \left(\frac{f(y)}{u(E)\Delta s(p_A, p_B)}\right)\right] + \left[p_A \left(\frac{\partial \Delta s}{\partial p_A}\right) \frac{yf(y)}{\Delta s(p_A, p_B)}\right]. \tag{2.9}$$
$$\pi_2^B(p_A, p_B) = [F(y)] - \left[p_B \left(\frac{f(y)}{u(E)\Delta s(p_A, p_B)}\right)\right] + \left[p_B \left(-\frac{\partial \Delta s}{\partial p_B}\right) \frac{yf(y)}{\Delta s(p_A, p_B)}\right]. \tag{2.10}$$

Numbered subscripts of *functions* represent partial derivatives with respect to the corresponding variable. (2.9) and (2.10) provide an intuitive understanding of the different effects of a marginal increase in provider pricing. The first term in each equation is the direct positive effect of revenue increases from the price increase. The second term is the direct negative effect that arises out of the reduction in revenues from the direct reduction in demand that accompanies the price increase. The final term, which we term the *externality effect*, captures the moderating effect of the negative usage externality. When a provider's price increases, this shifts demand from their network to that of their competitor, thereby changing the quality difference between the two providers in favor of the one whose price has increased. The externality effect captures the revenue impact of this change in quality difference. It is easily verified that this effect is strictly positive in both equations. Other things being equal, the presence of the externality therefore results in an equilibrium increase in prices, since it increases the slope of the second-period payoff functions of *both* providers.

3. Equilibrium

This section specifies the fulfilled-expectations subgame perfect Nash equilibria of the game introduced in Section 2. For ease of tractability, we assume that θ is uniformly distributed[9] in $[0, 1]$, with corresponding distribution function $F(\theta) = \theta$ and density function $f(\theta) = 1$. The derivation of the equilibrium has three parts. First, we formalize the fixed-point equation that determines fulfilled-expectations outcomes after infrastructure and prices have been chosen. Next, we specify the Nash equilibrium of the second-period price competition game. Finally, we specify the first-period subgame perfect Nash equilibrium choices of infrastructure.

3.1. Fulfilled-expectations outcomes

As described in Section 2.2, given a set of prices (p_A, p_B) and capacities (N_A, N_B), customers form an expectation of the value of y, the indifferent customer type (also referred to as the *market-share expectation*). Given y, there is a unique expectation of quality levels, as originally specified in

[9] This is simply equivalent to a rescaling of the type parameter θ; therefore, for results that are directional or comparative (rather than based on an interpretation of the absolute magnitude of some variable or function), this assumption is without loss of generality.

equations (2.2) and (2.3):

$$s_A(y) = 1 - B(v, \frac{E(1-y)}{N_A}). \quad (3.1)$$

$$s_B(y) = 1 - B(v, \frac{Ey}{N_B}). \quad (3.2)$$

Given an arbitrary y, N_A, N_B, define the *quality differential function* $Q(y, N_A, N_B)$ as the difference $[s_A(y) - s_B(y)]$. Using (3.1) and (3.2), this function is:

$$Q(y, N_A, N_B) = B(v, \frac{Ey}{N_B}) - B(v, \frac{E[1-y]}{N_A}). \quad (3.3)$$

The RHS of (3.3) has an expression identical to the function $\Delta s(p_A, p_B)$ used in Section 2.3. Given the expectations of quality, the *actual value* of the indifferent customer type, denoted $\Gamma(y)$, is specified below.

$$\text{When} \quad p_A > p_B : \Gamma(y) = \frac{p_A - p_B}{u(E)Q(y, N_A, N_B)}. \quad (3.4)$$

$$\text{When} \quad p_A = p_B : \Gamma(y) = \begin{cases} 0, & Q(y, N_A, N_B) < 0; \\ 1, & Q(y, N_A, N_B) > 0; \\ \frac{N_B}{N_A + N_B}, & Q(y, N_A, N_B) = 0. \end{cases} \quad (3.5)$$

If provider prices are different, then the indifferent customer type is as specified by (2.4). If prices are equal, but customer expectations are such that one provider has a higher quality, then that provider gets all the customers. On the other hand, if prices are equal, and if the market-share expectation y is such that service quality is equal, then the providers split demand based on network deployment levels. With equal prices and expected quality levels, any market share split is reasonable. (3.5) specifies the outcome which ensures that $\Gamma(y)$ is continuous[10].

Therefore, given a set of prices (p_A, p_B) and capacities (N_A, N_B), a market-share expectation y^*

[10] Often, the assumption made under this scenario is that firms split demand in any ratio. In fact, this may appear to be a critical assumption, since our equilibrium sometimes involves equal prices and quality. However, we will establish in Section 3.2 that with symmetric networks $N_A = N_B$, the only split in demand consistent with fulfilled expectations and second-period profit maximization is one in which providers split the market equally; moreover, with $N_A \neq N_B$, prices are never equal. Therefore, this assumption is without loss of generality.

satisfies *fulfilled-expectations* if it is a solution to

$$y^* = \Gamma(y^*). \tag{3.6}$$

3.2. Equilibrium prices

The following proposition establishes the conditions for a pair of prices (p_A, p_B) to be a Nash equilibrium that satisfies fulfilled-expectations. All proofs are in Appendix A.

Proposition 1. *Given N_A and N_B, any second-period Nash equilibrium in pure strategies takes the following form:*

$$p_A = (1-y)u(E)[Q(y, N_A, N_B) + y[1-y]u(E)Q_1(y, N_A, N_B)]; \tag{3.7}$$

$$p_B = yu(E)Q(y, N_A, N_B) + y^2u(E)Q_1(y, N_A, N_B)], \tag{3.8}$$

where $y \in [0,1]$ satisfies fulfilled expectations as specified by (A.6), given p_A, p_B and N_A, N_B

Proposition 1 specifies necessary (rather than sufficient) conditions for a second-stage Nash equilibrium that satisfies fulfilled expectations. The component of equilibrium price that depends on the direct difference in quality – that is, $(1-y)u(E)Q(y, N_A, N_B)$ for firm A and $yu(E)Q(y, N_A, N_B)$ for firm B – is referred to as the *quality premium*. These would be the equilibrium prices in the absence of usage externalities. The other component, which was discussed in section 2.3 – $u(E)(1-y)yQ_1(y, N_A, N_B)$ for firm A, and $u(E)y^2Q(y, N_A, N_B)$ for firm B – is referred to as the *externality premium*.

Proposition 2 establishes some properties of a symmetric second-period price equilibrium.

Proposition 2. *(a) If $N_A = N_B = N$, then there exists a unique symmetric second-period Nash equilibrium (p^*, p^*), where:*

$$p^* = \frac{Eu(E)}{2N} B_2(v, \frac{E}{2N}).$$

(b) If $N_A \neq N_B$, then there is no symmetric second-period Nash equilibrium in pure strategies,

and any asymmetric second-period Nash equilibrium in pure strategies takes the form specified by (3.7) and (3.8).

3.3. Equilibrium network infrastructure levels

This subsection shows that the subgame-perfect equilibrium network infrastructure deployment levels N_A^*, N_B^* are either at symmetric at the minimum required level $N_{\min}$ or are asymmetric. Given N_A, N_B, denote the corresponding equilibrium prices as $p_A(N_A, N_B)$ and $p_B(N_A, N_B)$ respectively, and the corresponding equilibrium indifferent customer as $y(N_A, N_B)$. From Proposition 1:

$$\begin{aligned} p_A(N_A, N_B) &= u(E)[1 - y(N_A, N_B)][Q(y(N_A, N_B), N_A, N_B) + y(N_A, N_B)Q_1(y(N_A, N_B), N_A, N_B)]; \\ p_B(N_A, N_B) &= u(E)y(N_A, N_B)[Q(y(N_A, N_B), N_A, N_B) + y(N_A, N_B)Q_1(y(N_A, N_B), N_A, N_B)], \end{aligned}$$

where $y(N_A, N_B)$ is implicitly defined by:

$$u(E)Q(y(N_A, N_B), N_A, N_B)y(N_A, N_B) = p_A(N_A, N_B) - p_B(N_A, N_B). \tag{3.9}$$

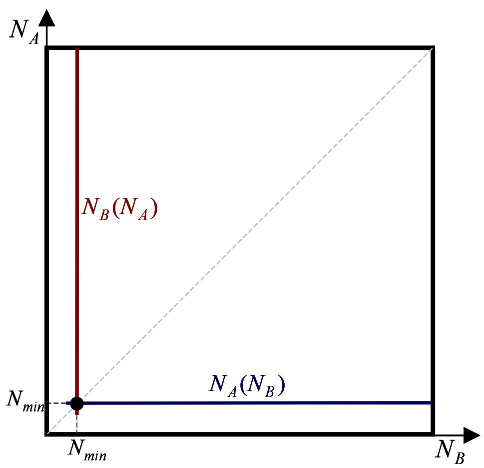
The equilibrium infrastructure levels of the providers can now be characterized:

Proposition 3. *Any subgame perfect Nash equilibrium of the deployment-pricing game takes one of the following forms:*

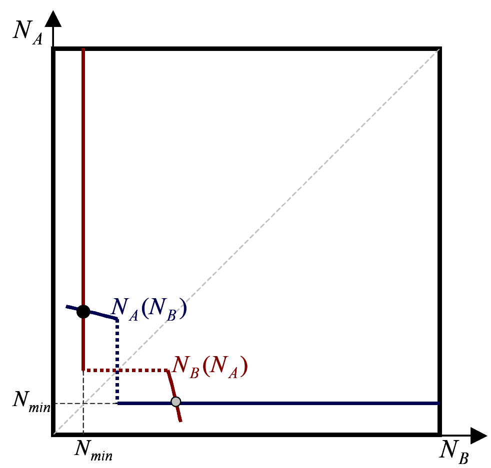
(a) Both providers deploy the minimum infrastructure, and price symmetrically: $N_A^* = N_B^* = N_{\min}$ *and* $p_A^* = p_B^* = \frac{Eu(E)}{2N_{\min}} B_2(v, \frac{E}{2N_{\min}})$

(b) Providers deploy asymmetric network capacity, and charge different prices. That is, if either $N_A^* > N_{\min}$ *or* $N_B^* > N_{\min}$, *then* $N_A^* \neq N_B^*$ *and* $p_A^* \neq p_B^*$.

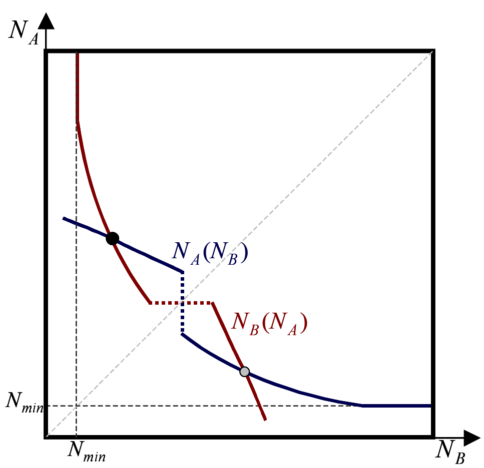
Since the providers have symmetric payoff functions, if (N_A^*, N_B^*) is an equilibrium pair, then so is (N_B^*, N_A^*). We continue to denote the higher quality provider as provider A. A closer analysis of the first-period subgame-perfect Nash equilibrium reveals that there are three specific kinds of equilibrium pairs, each associated with a distinct subset of the model's parameters, which are discussed



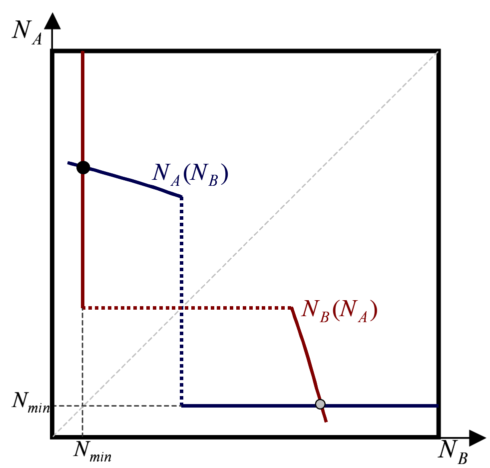
(a) Symmetric minimum infrastructure



(b) Asymmetric, one provider deploys $N_{\min}$.



(c) Interior solution with N_A, $N_B > N_{\min}$



(d) Asymmetric, one provider deploys $N_{\min}$.again

Figure 3.1: Depicts the best response functions for the two providers for different parameter values. The equilibrium points are circled. The solid back circle indicates the equilibrium point discussed in the text (where firm A is the higher quality provider), and the grey circle indicates the other equilibrium, which is identical except for a relabeling of providers. The discontinuity in the curves occurs when the firms 'switch positions' – that is, the point at which the deployment of provider j is high enough to make provider i's optimal response a choice which makes them the lower-quality, lower-capacity provider.

below for a fixed $c(N)$ and v, framed in terms of ranges of and increases in *average traffic* E. For convenience, the first-period payoff functions are reproduced below:

$$\Pi^A(N_A, N_B) = u(E)[1 - y(N_A, N_B)]^2[Q(y, N_A, N_B) + y(N_A, N_B)Q_1(y(N_A, N_B), N_A, N_B)] - c(N_A)$$

$$\Pi^B(N_A, N_B) = u(E)[y(N_A, N_B)]^2[Q(y, N_A, N_B) + y(N_A, N_B)Q_1(y(N_A, N_B), N_A, N_B)] - c(N_B)$$

There are two values $E = \underline{E}$ and $E = \overline{E}$, with $\underline{E} < \overline{E}$, that satisfy the first-order condition of the higher quality provider at the minimum deployment level $N_{\min}$:

$$\Pi_1^A(N_{\min}, N_{\min}) = 0 \tag{3.10}$$

Outside the range $(\underline{E}, \overline{E})$, both providers choose equilibrium strategies according to part (a) of Proposition 3, deployment is at the minimum level $N_{\min}$, and prices are equal. The best response functions and the corresponding equilibrium is depicted in Figure 3.1(a).

When E is higher than threshold level $\underline{E}$, it is no longer optimal for provider A to choose $N_{\min}$, since $\Pi_1^A(N_{\min}, N_{\min}) > 0$ for $E > \underline{E}$. Denote the best response of provider A to a choice of $N_{\min}$ by provider B as $N_{\min}^A(E)$:

$$N_{\min}^A(E) = \arg\max_N \Pi^A(N, N_{\min}). \tag{3.11}$$

For a range of values of $E > \underline{E}$, $N_{\min}^A(E)$ is the equilibrium choice of provider A, and provider B continues to deploy at the level $N_{\min}$. This is depicted in Figure 3.1(b), and remains the equilibrium until a second threshold value E', which is one of the two values of E that solve:

$$\Pi_1^B(N_{\min}^A(E), N_{\min}) = 0 \tag{3.12}$$

Since $\Pi_1^B(N_{\min}^A(E), N_{\min}) > 0$ for $E > E'$ in the neighborhood of E', both providers choose interior deployment levels $N_A^* > N_{\min}, N_B^* > N_{\min}$, as depicted in Figure 3.1(c).

At the next threshold value E'', which is the second value of E that solves (3.12), $N_B^* = N_{\min}$ again, and provider B chooses $N_{\min}$ for all $E > E''$. Provider A continues to increase their deployment

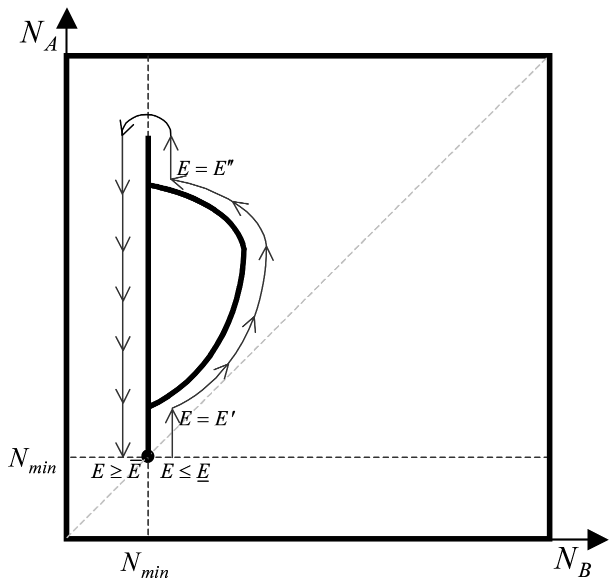


Figure 3.2: Evolution of the equilibrium infrastructure deployment point (first-stage choices of N) as average traffic per user E varies.

as E increases, yielding an equilibrium as depicted in Figure 3.1(d), then begins to reduce deployment until E reaches the threshold $\overline{E}$, the second value of E that solves (3.10). At this point and beyond, provider A also chooses $N_{\min}$, since $\Pi_1^A(N_{\min}, N_{\min}) < 0$ for $E > \overline{E}$, and the equilibrium returns to the one depicted in Figure 3.1(a).

Figure 3.2 summarizes the evolution of the first-stage equilibrium deployment of network infrastructure as average per-user traffic E varies. The outcomes are identical at extreme values of E. When $E < \underline{E}$, there is insufficient traffic to justify infrastructure investment beyond the minimum level $N_{\min}$. On the other hand, when $E > \overline{E}$, the traffic is sufficiently high that the marginal impact of an additional base station on service quality is too low to justify its costs.

4. Analysis and managerial implications

This section analyzes how revenues, profits, network infrastructure levels, service quality and market share vary with changes in average traffic E, spectrum v and infrastructure costs $c(N)$. This analysis leads to managerial prescriptions for network deployment, pricing and technology migration. Owing

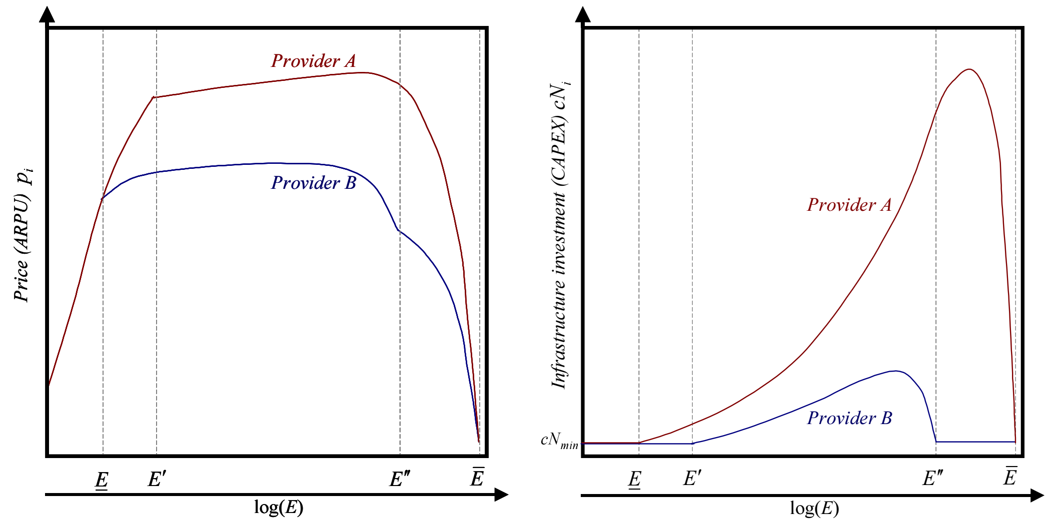


Figure 4.1: Variation in prices (or average revenue per user) and investements in network infrastructure (CAPEX) as average traffic E increases.

to the complexity of the payoff functions derived in section 3, and the non-monotonicity of the higher derivatives of the Erlang loss function $B(v, \rho)$, comparative statics results are difficult to obtain analytically. We have therefore computed the equilibria of the game for a number of specific parameter values, and obtained the corresponding economic measures. We summarize those results from this exercise that appeared consistently across our numerical analysis.

The analysis of this section assumes that infrastructure costs are linear (that is, $c(N) = cN$), and that $u(E) = E^\alpha$, where $\alpha \in [0, 1)$ measures the rate of increase in value as average per-user traffic increases.

4.1. Pricing strategy, network deployment and revenue trends

We partition the range of values of E into three sets: *early-stage* markets for which $E \leq E'$, *mature* markets for which $E' \leq E \leq E''$, and *declining* markets for which $E \geq E''$. The rationale for this partition will be clearer through the discussion that follows.

Figure 4.1 illustrates the variation in prices and network deployment for a candidate set of values of $\alpha > 0$. Each of the figures 4.1 through 4.4 uses exactly the same parameter values. If α is high enough, there is an initial increase in prices as E increases; however, at higher values of E, the quality

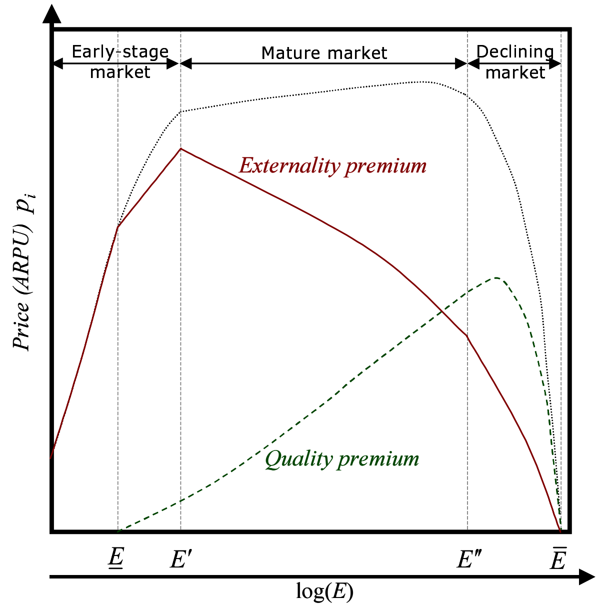


Figure 4.2: Illustrates the contribution of the quality premium and the externality premium to the price charged by Provider A as E varies. Also identifies the partition of the values of E into the early-stage, mature and declining market phases.

deterioration from the negative usage externalities dominates the value increases from an increase in $u(E)$, which results in a decrease in incremental infrastructure deployment and a corresponding decline in prices. In equilibrium, the providers initially offset the reduction in quality partially by increasing their expenditure on infrastructure, as also illustrated in Figure 4.1. This tradeoff leads to prices (ARPU levels) that are relatively flat in the mature market phase, which is consistent with observed industry trends.

Figure 4.2 provides a clearer illustration of the relative contribution of the two components of provider pricing – the *quality* premium and the *externality* premium – derived in (3.7), for Provider A. The externality premium is the primary driver of pricing power for lower levels of per-user traffic E. As average traffic increases, Provider A invests proportionately more than Provider B, increasing the quality gap and the quality premium. However, this rate of change eventually decreases as N_i increases, due to the fact that the rate of change of the quality differential $Q_1(y, N_A, N_B)$ is linearly increasing in both E/N_A and E/N_B. As a consequence, the externality premium declines. At a point beyond E'', Provider A's rate of infrastructure addition slows, the quality differential peaks, and then begins to fall, more rapidly as Provider A reduces their infrastructure investment.

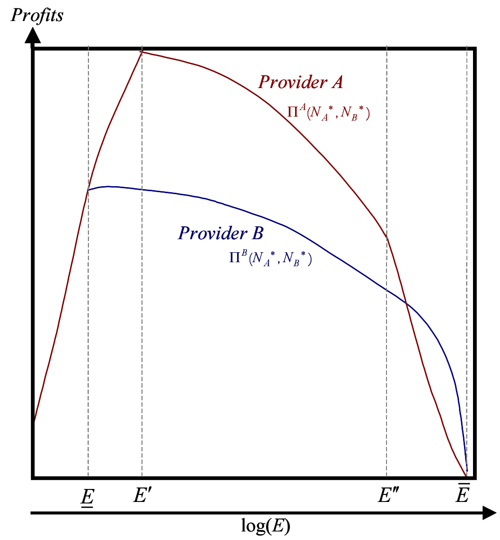


Figure 4.3: Variation in equilibrium profits (earnings) $\Pi^A(N_A^*, N_B^*)$ and $\Pi^B(N_A^*, N_B^*)$ as average traffic per user E increases.

The decomposition above leads to the following prescribed pricing and deployment strategy in the first two phases of a wireless market's evolution. In early-stage markets, when average traffic per user is low, providers should deploy relatively low levels of network infrastructure. Significant vertical differentiation is not necessary (or optimal) for either firm, since positive equilibrium prices can be maintained on account of the externality premiums. Additionally, substantial increases in quality will increase the price elasticity of demand (informally, because the effect of a marginal transfer of market share on quality levels decreases as quality increases), which is likely to intensify price competition. However, as the market matures and the average traffic per user increases significantly, the externality premium declines and the primary source of pricing power is from differentiation based on quality. Consequently, during this phase, firms need to pursue more aggressive quality differentiation, which will naturally be accompanied by rapid and continuous deployment of networking infrastructure.

4.2. Profits, quality and technology migration

The pricing and infrastructure deployment trends highlighted in Section 4.1 clearly indicate that provider profits will decrease as average traffic increases over the interval where revenues decrease while costs increase. This is indeed the case, as illustrated in Figure 4.3; while equilibrium profits increase with average traffic for Provider A in the early-market phase, they decline bilaterally in the mature market phase. More importantly, there is a steep bilateral decline in profits during the declining market phase. In fact, at $E = \overline{E}$, the earnings of both firms are consistently very close to zero, and this occurs across the entire range of parameter values we have studied. In other words, the point at which both providers stop investing in infrastructure (and choose equilibrium deployment $N_A^* = N_B^* = N_{\min}$) in response to increases in average per-user traffic coincides almost exactly with the point at which total provider profits are bilaterally zero.

This observation can be explained by the fact that the equilibrium traffic per base station is always fairly high at $E = \overline{E}$. Suppose one were to use the heavy-traffic approximation for the Erlang loss function:

$$B(v, \rho) \simeq \left(1 - \frac{v}{\rho}\right). \tag{4.1}$$

Under this approximation, the quality differential function is:

$$Q(y, N_A, N_B) = \frac{v}{E}\left[\frac{N_A}{1-y} - \frac{N_B}{y}\right], \tag{4.2}$$

and therefore,

$$yQ_1(y, N_A, N_B) = \frac{v}{E}\left[\frac{yN_A}{[1-y]^2} - \frac{N_B}{y}\right]. \tag{4.3}$$

As a consequence, provider A's profit function takes the form:

$$\begin{aligned}\pi^A(N_A, N_B) &= u(E)[1-y]^2[Q(y, N_A, N_B) + yQ_1(y, N_A, N_B)] && (4.4)\\ &= u(E)\frac{v}{E}N_A - cN_A. && (4.5)\end{aligned}$$

Recall that N_B^* remains fixed at$N_{\min}$ for values of average traffic $E' < E \leq \overline{E}$. From equation (4.5), it is clear that under this approximation, the same value of N_A solves both $\Pi^A(N_A, N_{\min}) = 0$ and $\Pi_1^A(N_A, N_{\min}) = 0$. Therefore, the point at which provider A stops investing in infrastructure is the same as the point at which their profits are zero. Intuitively, under heavy traffic, each part of the provider's network infrastructure is operating at close to full utilization (the approximation in (4.1) is analytically identical to the assumption of 100% utilization). The marginal revenue from

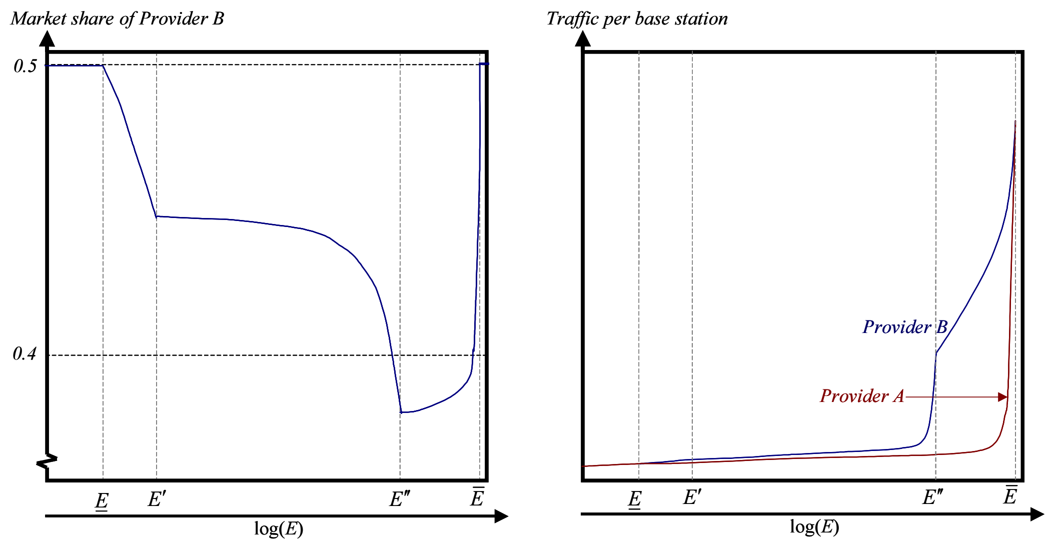


Figure 4.4: Variation in market share and average traffic per base station as average demand per user E increases.

each unit of infrastructure (that is, each base station) is therefore approximately equal, since each reduces the loss rate by about the same fraction $\frac{v}{E(1-y)}$. Consequently, provider A's profit function is approximately linear in N_A, and when marginal revenue from an additional base station reaches the variable cost of the base station c, profits tend to zero.

The variation in relative market share and base station load, illustrated in Figure 4.4, indicate that in the mature market phase (that is, beyond the threshold E''), there is a steep decline in service quality as well. This is due to the exponential increase in average traffic per base station that follows from the simultaneous increase in average traffic and lowering of equilibrium network infrastructure.

A logical response to this decline in quality and profits is to 'shift' the equilibrium by changing the number of effective channels v available to each provider. As illustrated partially in Figure 4.5, an increase in v does not change the qualitative features of the profit, price or network infrastructure deployment trends, but merely changes the range of values of E over which they occur. Specifically, an increase in v 'stretches' out the profit curves, while a decrease tends to 'compress' them. This is because the threshold values of E – that is, the values of $\underline{E}$, E', E'', and $\overline{E}$ defined in section 3.4 – are all increasing in v, the number of effective channels, which is substantially higher with each new

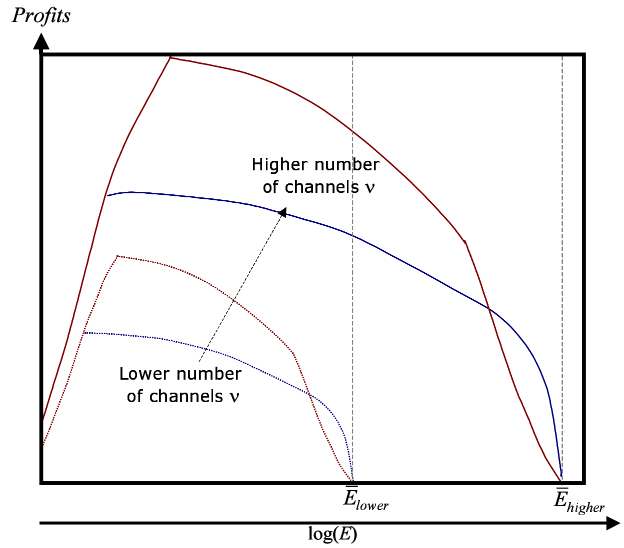


Figure 4.5: Impact of migrating to a new generation of technology. An increase in the number of effective channels v increases earnings bilaterally, and also increases the range of demand over which positive earnings are sustainable in equilibrium.

generation of transmission technology[11].

The discussion above indicates that the declining market phase $E \in [E'', \overline{E}]$ represents a natural range of traffic over which providers should actively execute their migration to a new generation of wireless technology. Under the existing generation of transmission technology, both prices and profits decline sharply with increases in customer usage, and equilibrium quality levels decline to a point where the market is no longer likely to be viable. As E reaches the upper threshold $\overline{E}$, equilibrium profits are zero, which makes technology migration a necessity for continued business survival. Switching to a technology that admits a higher number of effective channels will bilaterally increase earnings, as depicted by Figure 4.5, and this increase sustains even if the costs of infrastructure for the new technology are higher.

Two other interesting observations that emerged from our analysis:

1. In a corresponding model *without* externalities and with a direct choice of s_A and s_B, it is

[11] The impact of a decrease in per-base-station infrastructure costs c are similar. In addition to the corresponding changes in the threshold values of E (each of which is decreasing in c), there is also a direct positive effect on deployment at each value of E, since the (decreasing) marginal revenue from infrastructure deployment is equated to a lower marginal fixed cost c.

straightforward to show that independent of c, firm B would have a market share of $\frac{1}{3}$, and firm A would have a market share of $\frac{2}{3}$, a more uneven division of market share than illustrated in Figure 4.4. The presence of usage externalities explains the more symmetric equilibrium division of market share, which is a consequence of their moderating a firm's need to (and ability to profitably) differentiate their product.

2. In the declining market phase, the profits of the higher quality provider A are sometimes lower than those of the lower quality provider B. This occurs when the infrastructure deployment of provider B is constant at $N_{\min}$, and when provider A, while having higher demand, faces disproportionately higher equilibrium infrastructure costs. Since this occurs in a range of traffic over which we anticipate active technology migration, we do not interpret it further.

5. Summary and directions for future work

Our model of competition in the wireless telecommunications industry captures the interplay between network deployment, the utilization of spectrum by the generation of transmission technology, the negative loss-rate based externalities that reduce service quality as demand increases, and the presence of required minimum network infrastructure levels driven by technological and economic constraints. While our model has been developed to study the wireless industry, it extends a standard model of quality-differentiated competition to admit loss-rate based externalities and minimum capacity constraints, while preserving endogenous sequential capacity and pricing choices, and we hope it will apply to other IS economics questions of interest. The key managerial insights from our analysis are summarized below.

– The intensity of competition in the wireless industry is mediated by the presence of usage externalities which increase the pricing power of competing providers, especially in early-stage markets. At this stage, providers should therefore ideally maintain minimum infrastructure levels and avoid active quality differentiation.

– As wireless markets mature, externality premiums decline steadily, and providers need to pursue an aggressive quality-based differentiation strategy in order to maintain their pricing power. This will require continuous and rapid growth of network infrastructure. During this phase of competition, relatively flat average revenue per user is a natural equilibrium outcome, even though usage levels and the value of the wireless services offered by both firms may be rising steadily.

– There is a critical level of average per-user traffic at which viable service quality levels and positive profits are not sustainable. This is preceded by a declining market phase, over which profits will fall sharply if a provider continues to utilize their existing spectrum with their existing generation of technology. Consequently, the onset of this phase should trigger active migration to the next generation of transmission technology, generally accompanied by a shift to a new block of spectrum and investment in new network assets.

As new bandwidth-intensive services continue to drive up the average traffic per user, these newer networks will reach their corresponding critical $\overline{E}$ values, and the cycle of technology migration is likely to persist. In this context, if calibrated and extended to a market with multiple providers, our model might help regulators price future spectrum and predict equilibrium industry structure. This is one direction for future research that seems promising. Since our model demonstrates higher pricing power in the industry due to the presence of externalities, this extension may also form a basis for adjusting industry-concentration based measures like the Herfindahl-Hirschman index, which are likely to underestimate market power in the industry. This is likely to be of particular interest if consolidation in the industry continues.

Another logical direction of future research is to explicitly study a technology migration game that uses a simplified version of our model as a stage game. A limitation of our model is its interpretation of a dynamic phenomenon using comparative statics, and this would be a suitable extension to permit dynamic forward-looking infrastructure choice; it may also indicate whether the timing of migration to next-generation technologies occurs at a point which is too early or too late, and whether this outcome can be altered by appropriately timed government subsidies.

6. References

1. Armony M. and M. Haviv, 2003, "Price and Delay Competition between Two Service Providers," *The European Journal of Operational Research*, 147 (1), 32-50.

2. Asvanund, A., Clay, C. Krishnan, R. and Smith, M., 2004, "An Empirical Analysis of Network Externalities in Peer-to-Peer Music-Sharing Networks," *Information Systems Research* 15, 155-174.

3. Ausubel, L., P. Cramton, P. McAfee and J. McMillan, 1997, "Synergies in Wireless Telephony: Evidence from the Broadband PCS Auctions" *Journal of Economics and Management Strategy*, 6 (3), 497-527.

4. Bhargava, H. and Choudhary, V., 2004, "Economics of an Information Intermediary with Aggregation Benefits," *Information Systems Research* 15, 22-36.

5. Dewan, S. and Mendelson, H., 1990, "User Delay Costs and Internal Pricing for a Service Facility," *Management Science* 36, 1502-1517.

6. Hosanagar, K., Krishnan, R., Chuang, J. and Choudhary, V., 2003, "Pricing and Resource Allocation in Caching Services with Multiple Levels of QoS," available at *http://oz.stern.nyu.edu/seminar/sp03/0205.pdf*

7. Jagerman, D., Melamed, B. and Willinger, W., 1996, "Stochastic Modeling of Traffic Processes," Available at *http://citeseer.ist.psu.edu/jagerman96stochastic.html*

8. Laffont, J-J, P. Rey and J. Tirole, 1998, "Network Competition I. Overview and Nondiscriminatory Pricing," *Rand Journal of Economics*, 29 (1), 1-37

9. Lederer, P. J. and L. Li, 1997, "Pricing, Production, Scheduling, and Delivery-Time Competition," *Operations Research* 45 (3), 407-420.

10. Levhari, D. and I. Luski, 1978, "Duopoly Pricing and Waiting Lines," *European Economic Review*, 11, 17-35.

11. Luski, I., 1976, "On Partial Equilibrium in a Queuing System with Two Servers," *Review of Economic Studies*, 43, 519-525.

12. Mendelson, H., 1985, "Pricing Computer Services: Queuing Effects," *Communications of the ACM* 28, 312-321.

13. Mendelson, H. and Shneorson, S., 2003, "Internet Peering, Capacity and Pricing,", Working Paper, Stanford University.

14. Mendelson, H. and Whang, S., 1990, "Optimal Incentive-compatible Priority Pricing for the $M/M/1$ Queue, " *Operations Research* 38, 870-883.

15. Nadiminti, R., Mukhopadhyay, T. and Kriebel, C., 2002, "Intrafirm Resource Allocation with Asymmetric Information and Negative Externalities," *Information Systems Research* 13, 428-434.

16. Reiffen D., L. Schumann and M. Ward, 2000, "Discriminatory Dealing with Downstream Competitors: Evidence from the Cellular Industry," *Journal of Industrial Economics* XLVIII, 253-286.

17. Reitman, D., 1991, "Endogenous Quality Differentiation in Congested Markets," *The Journal of Industrial Economics*, 39 (6), 621-647.

18. Shaked, A. and J. Sutton, 1982, "Relaxing Price Competition through Product Differentiation," *Review of Economics Studies*, XLIX, 3-13.

19. Spulber, D., 1995, "Introduction: Competition in Telecommunications," *Journal of Economics and Management Strategy,* 4 (2), 133-137.

20. Sweet, R., I. Viehoff, D. Linardatos, and N. Kalouptsidis, 2001, "Marginal Value-based Pricing of Additional Spectrum Assigned to Cellular Telephony Operators," *Information Economics and Policy* 14 (3), 371-384.

21. Valetti, T., 1999, "A Model of Competition in Mobile Communications," *Information Economics and Policy* 11, 61-72.

22. Westland, C., 1992, "Congestion and Network Externalities in the Short-run Pricing of Information Systems Services," *Management Science* 38, 992-1099.

A. Proofs

Proof of Proposition 1: The second-period payoffs to each provider are:

$$\pi^A(p_A, p_B) = p_A[1-y], \tag{A.1}$$

$$\pi^B(p_A, p_B) = p_B y, \tag{A.2}$$

where y satisfies fulfilled expectations. Any Nash equilibrium price pair must satisfy first-order conditions $\pi_1^A(p_A, p_B) = \pi_2^B(p_A, p_B) = 0$. Differentiating (A.1) with respect to p_A yields:

$$1 - y - p_A \frac{\partial y}{\partial p_A} = 0, \tag{A.3}$$

which rearranges to:

$$p_A = \frac{1-y}{\frac{\partial y}{\partial p_A}}. \tag{A.4}$$

Similarly, differentiating (A.2) with respect to p_B and rearranging yields:

$$p_B = \frac{-y}{\frac{\partial y}{\partial p_B}}. \tag{A.5}$$

Based on (3.4) and (3.6), the condition for y to satisfy fulfilled expectations can be summarized as:

$$yu(E)Q(y, N_A, N_B) = p_A - p_B. \tag{A.6}$$

Since y satisfies fulfilled-expectations, we can differentiate both sides of (A.6) with respect to p_A to get:

$$\frac{\partial y}{\partial p_A} u(E) \left(B(v, \frac{Ey}{N_B}) - B(v, \frac{E(1-y)}{N_A}) \right) + yEu(E) \frac{\partial y}{\partial p_A} \left(\frac{1}{N_B} B_2(v, \frac{Ey}{N_B}) + \frac{1}{N_A} B_2(v, \frac{E(1-y)}{N_A}) \right) = 1, \tag{A.7}$$

which implies that

$$\frac{\partial y}{\partial p_A} = \frac{1}{u(E) \left[B(v, \frac{Ey}{N_B}) - B(v, \frac{E[1-y]}{N_A}) + yE \left(\frac{1}{N_B} B_2(v, \frac{Ey}{N_B}) + \frac{1}{N_A} B_2(v, \frac{E(1-y)}{N_A}) \right) \right]}. \tag{A.8}$$

Similarly, differentiating both sides of (A.2) with respect to p_B yields:

$$\frac{\partial y}{\partial p_B} u(E) \left(B(v, \frac{Ey}{N_B}) - B(v, \frac{E(1-y)}{N_A}) \right) + yEu(E) \frac{\partial y}{\partial p_B} \left(\frac{1}{N_B} B_2(v, \frac{Ey}{N_B}) + \frac{1}{N_A} B_2(v, \frac{E(1-y)}{N_A}) \right) = -1, \tag{A.9}$$

which implies that

$$\frac{\partial y}{\partial p_B} = \frac{-1}{u(E) \left[B(v, \frac{Ey}{N_B}) - B(v, \frac{E[1-y]}{N_A}) + yE \left(\frac{1}{N_B} B_2(v, \frac{Ey}{N_B}) + \frac{1}{N_A} B_2(v, \frac{E(1-y)}{N_A}) \right) \right]}. \tag{A.10}$$

Now, differentiating $Q(y, N_A, N_B)$ with respect to y yields:

$$Q_1(y, N_A, N_B) = E \left(\frac{1}{N_A} B_2(v, \frac{E[1-y]}{N_A}) + \frac{1}{N_B} B_2(v, \frac{Ey}{N_B}) \right) \tag{A.11}$$

Comparing (A.8) and (A.10) with (A.11) and (3.3) yields:

$$\frac{\partial y}{\partial p_A} = \frac{1}{u(E)[Q(y, N_A, N_B) + yQ_1(y, N_A, N_B)]}, \tag{A.12}$$

and

$$\frac{\partial y}{\partial p_B} = \frac{-1}{u(E)[Q(y, N_A, N_B) + yQ_1(y, N_A, N_B)]}. \tag{A.13}$$

The result follows by substituting (A.12) into (A.4) and (A.13) into (A.5).

Proof of Proposition 2: From (3.7) and (3.8) in Lemma 1, any equilibrium price pair (p_A, p_B) must satisfy

$$p_A = [1-y]u(E)[Q(y, N_A, N_B) + yQ_1(y, N_A, N_B)]. \tag{A.14}$$

$$p_B = yu(E)[Q(y, N_A, N_B) + yQ_1(y, N_A, N_B)]. \tag{A.15}$$

Therefore,

$$p_A - p_B = (1-2y)u(E)[Q(y, N_A, N_B) + yQ_1(y, N_A, N_B)]. \tag{A.16}$$

It is straightforward to establish that $(0, 0)$ is not an equilibrium, and therefore, $[Q(y, N_A, N_B) + yQ_1(y, N_A, N_B)] > 0$. As a consequence, since $u(E) > 0$, (A.16) implies that if $p_A = p_B$, then $y = \frac{1}{2}$.

(a) If $N_A = N_B = N$, then (3.5) reduces to

$$\Gamma(y) = \frac{N}{2N}, \text{ if } B(v, \frac{Ey}{N}) = B(v, \frac{E(1-y)}{N}). \tag{A.17}$$

From the properties of the Erlang loss function $B(v, \rho)$, we know that $B_2(v, \rho) > 0$, which in turn implies that the unique value of y for which $B(v, \frac{Ey}{N}) = B(v, \frac{E(1-y)}{N})$ is specified by $\frac{Ey}{N} = \frac{E(1-y)}{N}$, or $y = \frac{1}{2}$. This implies that when $N_A = N_B = N$, (3.5) is exactly the same as:

$$\Gamma(y) = \frac{1}{2}, \text{ if } y = \frac{1}{2}. \tag{A.18}$$

Therefore, (A.18) is consistent with a value of y such that $\Gamma(y) = y$. Substituting $N_A = N_B = N$ and $y = \frac{1}{2}$ into (3.7) and (3.8) yields the expression for p^*. It can easily be established that $\pi^A(x, p^*)$ and $\pi^B(p^*, x)$ are each strictly quasiconcave in x, and therefore, (p^*, p^*) is a Nash equilibrium.

(b) When $N_A \neq N_B$, there is no combination of N_A, N_B that can ensure that $\Gamma(\frac{1}{2}) = \frac{1}{2}$ if $p_A = p_B$. As a consequence, no symmetric price pair can be a Nash equilibrium.

Proof of Proposition 3: Recall that we have assumed $p_A \geq p_B$, and therefore, for an indifferent customer y, firm A has market share $(1-y)$ and firm B has market share y. Also recall that the quality differential function, for any arbitrary y, N_A, N_B is:

$$Q(y, N_A, N_B) = B(v, \frac{Ey}{N_B}) - B(v, \frac{E[1-y]}{N_A}). \tag{A.19}$$

For reference, some partial derivatives of Q are listed below:

$$\begin{aligned}
Q_1(y, N_A, N_B) &= E\left(\frac{B_2(v, \frac{E[1-y]}{N_A})}{N_A} + \frac{B_2(v, \frac{Ey}{N_B})}{N_B}\right) \\
Q_2(y, N_A, N_B) &= \frac{E[1-y]}{N_A^2} B_2(v, \frac{E[1-y]}{N_A}) \\
Q_3(y, N_A, N_B) &= \frac{-Ey}{N_B^2} B_2(v, \frac{Ey}{N_B}) \\
Q_{11}(y, N_A, N_B) &= E^2\left(\frac{B_{22}(v, \frac{Ey}{N_B})}{N_B^2} - \frac{B_{22}(v, \frac{E[1-y]}{N_A})}{N_A^2}\right) \\
Q_{12}(y, N_A, N_B) &= -\left(\frac{E}{N_A^2} B_2(v, \frac{E[1-y]}{N_A}) + \frac{E^2[1-y]}{N_A^3} B_{22}(v, \frac{E[1-y]}{N_A})\right) \\
Q_{13}(y, N_A, N_B) &= -\left(\frac{E}{N_B^2} B_2(v, \frac{Ey}{N_B}) + \frac{E^2 y}{N_B^3} B_{22}(v, \frac{Ey}{N_B})\right)
\end{aligned}$$

The equilibrium second-period price pairs take the following form, as shown in Proposition 1:

$$\begin{aligned}
p_A(N_A, N_B) &= u(E)[1 - y(N_A, N_B)][Q(y(N_A, N_B), N_A, N_B) \\
&\quad + y(N_A, N_B) Q_1(y(N_A, N_B), N_A, N_B)] \qquad (A.20) \\
p_B(N_A, N_B) &= u(E) y(N_A, N_B)[Q(y(N_A, N_B), N_A, N_B) \\
&\quad + y(N_A, N_B) Q_1(y(N_A, N_B), N_A, N_B)] \qquad (A.21)
\end{aligned}$$

where $y(N_A, N_B)$ satisfies $\Gamma(y) = y$, which implies that $y(N_A, N_B)$ satisfies:

$$y = [1 - 2y][\frac{Q(y, N_A, N_B) + yQ_1(y, N_A, N_B)}{Q(y, N_A, N_B)}], \tag{A.22}$$

or rearranging, implies that $y(N_A, N_B)$ satisfies the relationship:

$$[3y(N_A, N_B) - 1]Q(y(N_A, N_B), N_A, N_B) = [1 - 2y(N_A, N_B)]y(N_A, N_B)Q_1(y(N_A, N_B), N_A, N_B). \tag{A.23}$$

Differentiating both sides of (A.23) successively with respect to N_A and $N_B$and rearranging yields:

$$y_1(N_A,N_B) = \frac{[1-3y(N_A,N_B)]Q_1(y(N_A,N_B),N_A,N_B)+y(N_A,N_B)[1-2y(N_A,N_B)]Q_{12}(y(N_A,N_B),N_A,N_B)}{3Q(y(N_A,N_B),N_A,N_B)+[7y(N_A,N_B)-2]Q_1(y(N_A,N_B),N_A,N_B)-[1-2y(N_A,N_B)]y(N_A,N_B)Q_{11}(y(N_A,N_B),N_A,N_B)} \tag{A.24}$$

$$y_2(N_A,N_B) = \frac{[1-3y(N_A,N_B)]Q_1(y(N_A,N_B),N_A,N_B)+y(N_A,N_B)[1-2y(N_A,N_B)]Q_{13}(y(N_A,N_B),N_A,N_B)}{3Q(y(N_A,N_B),N_A,N_B)+[7y(N_A,N_B)-2]Q_1(y(N_A,N_B),N_A,N_B)-[1-2y(N_A,N_B)]y(N_A,N_B)Q_{11}(y(N_A,N_B),N_A,N_B)} \tag{A.25}$$

The payoff functions of the two firms are

$$\Pi^A(N_A,N_B) = u(E)[1-y(N_A,N_B)]^2[Q(y,N_A,N_B) \tag{A.26}$$
$$+y(N_A,N_B)Q_1(y(N_A,N_B),N_A,N_B)] - c(N_A)$$
$$\Pi^B(N_A,N_B) = u(E)[y(N_A,N_B)]^2[Q(y,N_A,N_B) \tag{A.27}$$
$$+y(N_A,N_B)Q_1(y(N_A,N_B),N_A,N_B)] - c(N_B)$$

The first-order conditions are $\Pi_1^A(N_A,N_B) = \Pi_2^B(N_A,N_B) = 0$. *(For the remainder of this proof, we drop the variables in the expressions for brevity. For instance, we refer to $y_1(N_A,N_B)$ as y_1, $Q_{11}(y(N_A,N_B),N_A,N_B)$ as Q_{11} and so forth).*

Differentiating (A.26) with respect to N_A and (A.27) with respect to N_B yields:

$$\Pi_1^A = u(E)\{(1-y)^2(Q_2+yQ_{12}) + y_1(1-y)^2(2Q_1+yQ_{11}) \tag{A.28}$$
$$-2y_1(1-y)(Q+yQ_1)\} - c_1; \tag{A.29}$$
$$\Pi_2^B = u(E)\{y^2(Q_3+yQ_{13}) + y^2(y_2)(2Q_1+yQ_{11}) + 2y(y_2)(Q+yQ_1)\} - c_1, \tag{A.30}$$

which when rearranged, yield the first-order conditions:

$$(1-y)^2(Q_2+yQ_{12}) + y_1\left[y(1-y)^2Q_{11} + 2y(1-y)(1-2y)Q_1 - 2(1-y)Q)\right] = \frac{c_1}{u(E)} \tag{A.31}$$

$$y^2(Q_3+yQ_{13}) + y_2\left[y^3Q_{11} + 4y^2Q_1 + 2yQ)\right] = \frac{c_1}{u(E)} \tag{A.32}$$

Recall from (A.24) and (A.25) that

$$y_1 = \frac{(1-3y)Q_1 + y(1-2y)Q_{12}}{3Q+(7y-2)Q_1 - y(1-2y)Q_{11}} \tag{A.33}$$
$$y_2 = \frac{(1-3y)Q_1 + y(1-2y)Q_{13}}{3Q+(7y-2)Q_1 - y(1-2y)Q_{11}} \tag{A.34}$$

and therefore we have the first order conditions purely in terms of our primitive functions:

$$(1-y)^2(Q_2+yQ_{12})+\frac{[(1-3y)Q_1+y(1-2y)Q_{12}]\left[y(1-y)^2Q_{11}+2y(1-y)(1-2y)Q_1-2(1-y)Q)\right]}{3Q+(7y-2)Q_1-y(1-2y)Q_{11}}=\frac{c_1}{u(E)} \quad \text{(A.35)}$$

$$y^2(Q_3+yQ_{13})+\frac{[(1-3y)Q_1+y(1-2y)Q_{13}]\left[y^3Q_{11}+4y^2Q_1+2yQ)\right]}{3Q+(7y-2)Q_1-y(1-2y)Q_{11}}=\frac{c_1}{u(E)} \quad \text{(A.36)}$$

If $N_A = N_B = N$, this implies that $p_A = p_B = p$ (from Proposition 1) and that $y = \frac{1}{2}$. It therefore follows that:

$$Q_1 = \frac{2E}{N}B_2(v,\frac{E}{2N}) \quad \text{(A.37)}$$

$$Q_2 = -Q_3 = \frac{E}{2N^2}B_2(v,\frac{E}{2N}) \quad \text{(A.38)}$$

$$Q_{11} = 0 \quad \text{(A.39)}$$

$$Q_{12} = Q_{13} = -\left(\frac{E}{N^2}B_2(v,\frac{E}{2N})+\frac{E^2}{2N^3}B_{22}(v,\frac{E}{2N})\right) \quad \text{(A.40)}$$

and therefore:

$$\pi_1^A(N,N) = \frac{1}{4}(Q_2+\frac{Q_{12}}{2})-c_1 \quad \text{(A.41)}$$

$$\pi_2^B(N,N) = \frac{1}{4}(Q_3+\frac{Q_{13}}{2})-\frac{Q_1}{3}-c_1, \quad \text{(A.42)}$$

which upon substituting (A.38) and (A.40) yields:

$$\Pi_1^A(N,N) = -\frac{E^2}{16N^3}B_{22}(v,\frac{E}{2N})-c_1(N) \quad \text{(A.43)}$$

$$\Pi_2^B(N,N) = -\frac{E^2}{16N^3}B_{22}(v,\frac{E}{2N})-\left(\frac{E}{4N^2}+\frac{2E}{3N}\right)B_2(v,\frac{E}{2N})-c_1(N). \quad \text{(A.44)}$$

Since $B_2(v,\rho) > 0$ for all v,ρ, the two first-order conditions cannot be satisfied simultaneously, and the result follows.