Mobile termination rates and the receiver-pays regime

Ángel Luis López[1]

Public–Private Sector Research Center, IESE Business School

December 2008

[1]IESE Business School, Avda. Pearson, 21, 08034 Barcelona, Spain. Tel.: + 34 9 3253 4200; fax: + 34 9 3253 4343. E-mail address: alopezr@iese.edu. This article is a revised version of chapter 2 of my doctoral dissertation at the University of Toulouse 1. I am grateful to Patrick Rey for his guidance and helpful advice. I would also like to thank the participants at various seminars and conferences where I presented a previous version of this article. All remaining errors are my own.

Abstract

The European Commission has recently invited national regulatory authorities to decrease access charges to the cost of an efficient operator. Some large operators warned regulators and users that cutting access charges could result in the US style business model, where mobile users pay for both making and receiving calls. I show that mobile operators only find it profitable to charge for incoming calls when the access charge is below cost. In such a case profits are neutral with respect to the level of the access charge even if network operators compete in a multi-period model. I further show that 'bill and keep' may encourage consumers to make greater use of their mobiles. Finally, I discuss the policy implications of these results.

Keywords: Bill and Keep; Call externality; Mobile Telephony; Receiver pays; Termination rates

JEL classification: L41; L51; L96.

1 Introduction

In many European countries mobile network operators do not charge subscribers for receiving calls, whereas in the United States they usually charge their subscribers for the calls they receive. The first question that this article seeks to answer is why the liberalization of telecommunications has lead to these two different forms of competition for retail customers. A key finding is that it may be an endogenous price response to the level of the access charge. Access charges are those charges that mobile operators levy on each other and on fixed-line operators for terminating calls on their networks.

Access charges affect the cost of the off-net calls (i.e., those calls originated on a network and completed on a different network) and as a result have an impact on retail competition. This fact has raised two concerns among practitioners and regulators: the first is that network operators may use cooperation over interconnection to soften downstream competition[1], the second is that established network operators can use access charges to foreclose the market[2]. A third concern, which is the focus of this article, has recently raised: access charges may affect the form of competition for retail customers by inducing firms to adopt the caller-pays regime (only the caller pays for a call) or the receiver-pays regime (mobile subscribers also incur charges for receiving calls).

In contrast to Europe, in the United States access charges are quite likely below the cost of mobile termination, which may induce network operators to charge their customers for receiving calls so as to recover their cost. Indeed, this article shows that network operators only find it profitable to charge for incoming calls when the access charge is below cost. Additionally, this article examines the impact of the access charge on the equilibrium profit and social welfare when network operators adopt the receiver-pays regime.

Despite the spectacular growth of mobile telephony in the last years, access charges (for mobile-to-mobile and fixed-to-mobile calls) remained high in Europe. The reason is that call termination constitutes a competitive bottleneck: a mobile network operator always holds a monopoly over delivering calls to its customers and thus has an incentive to set a high access charge for providing call termination. The high access charges became a serious concern in most European countries. For example, in 2003 the UK Competition Commission concluded that high access charges operate against the public interest and that they were 30 to 40 per cent in excess of the estimated fair charge (Competition Commission, 2003).

[1] A large part of the literature on access charges and network competition, starting with the seminal work of Armstrong (1998) and Laffont et al. (1998a,b), has studied how access charges affect competition between networks. See, for instance, Armstrong and Wright (2008), Carter and Wright (2003), De Bijl and Peitz (2002, 2004), Dessein (2003), Gans and King (2000, 2001), Hahn (2004), Jeon and Hurkens (2008), López (2008a) and Valletti and Cambini (2005). See Armstrong (2002) for an excellent survey.

[2] See Calzada and Valletti (2008), Hoernig (2007) and López and Rey (2008).

Ex-ante regulation with price control obligations was imposed in Europe. Although these regulatory measures ordered mobile operators to cut access charges, these charges are still disproportionate to the real cost of mobile termination, actually call termination is still an important source of revenue. In particular, approximately 15% of the annual revenue for the sector in 2007 comes from mobile call termination; this "extra" profit has been an important means of recovering costs.

The European Commission has recently invited national regulatory authorities to set reciprocal (i.e., symmetric or uniform) access charges[3] and has also emphasized that access charges should be decreased to *the cost of an efficient operator* as soon as possible.[4] Similarly, in the Common Position adopted on February 2008,[5] the European regulators conclude that "economic principles tend to recommend the setting up of a unique and uniform termination rate for all network operators, determined with reference to costs incurred by an hypothetical efficient operator" as this will encourage productive efficiency.

If national regulatory authorities act according to the Commission and European regulators view and set a reciprocal access charge that equals the cost of an efficient operator, Will this affect the form of competition for retail customers? For example, in June 2008 the European Commissioner announced steps to cut termination rates (i.e., access charges) by up to 70%. Then, some large operators as for instance Vodafone, warned the European Union that cutting termination rates could mean the end of handset subsidies for consumers and raise retail charges. Actually, Vodafone said that cutting termination rates could result in US style business model, where mobile users pay for both making and receiving calls.

This article aims to explore this issue more fully. I examine networks operators' pricing strategies in a model in which customers derive utility from receiving calls (call externality), networks may charge for receiving calls and both the caller and the receiver can affect the length of a call by hanging up. Most of the literature on the impact of access charges on network competition has not given enough attention to these facts. Berger (2004, 2005) studies competition in the presence of termination-based price discrimination and call externalities,[6] however network operators are not allowed to charge for incoming calls. Jeon et al. (2004) and Laffont et al. (2003) are the most related articles to the concern that I address. Laffont et al. analyze Internet backbone competition and assumes that there exist two types

[3]Small operators usually benefit from higher access charges than large operators, which is justified by objective cost differences.

[4]The European Commissioner described these charges as "guaranteed money" for the mobile operators that moreover creates "real distortion of competition" (Financial Times, June 2008.)

[5]See "ERG's Common Position on symmetry of fixed call termination rates and symmetry of mobile call termination rates", adopted by the ERG-Plenary on 28th February 2008, p. 4-5. Available at http://erg.eu.int.

[6]See also Fabrizi (2005), Hermalin and Katz (2001, 2004) and Kim and Lim (2001).

of customers: senders or websites and receivers or consumers. In Jeon et al. (2004) and this article, however, every consumer both sends and receives traffic, and moreover obtains surplus from and is charged for making and receiving calls.

More precisely, Jeon et al. assume that the receiver's utility is subject to a random noise and a certain proportionality between the receiver's and the caller's utilities. They establish the existence (not uniqueness) of the 'off-net-cost pricing' equilibrium, which is described below, in the specific case in which the noise vanishes.[7] I generalize their framework by allowing a random noise in both the callers' and receivers' utilities, and by removing the assumption of a given proportionality between the utility functions. Yet the main contributions of this article are to show that the 'off-net-cost pricing' principle is a candidate equilibrium and, more importantly, to prove that under general conditions, which do not seem to be too restrictive, it exists and is the unique possible equilibrium. This result allows me to perform a policy-relevant comparative statics analysis: I show that in the 'off-net-cost pricing' equilibrium, the operators' profits are neutral with respect to the level of the access charge when they compete in either a static or multi-period model. Also, I derive the access charge that maximizes the *average* social welfare of a given call. By so doing, I show that 'bill and keep' encourages consumption and may be socially optimal. Finally, I discuss the policy implications of these findings.

The article is organized as follows. Section 2 describes the model. Section 3 examines retail competition for a given, reciprocal, access charge. It first characterizes the 'off-net-cost pricing' equilibrium and derives the profit-neutrality result, it then proves the existence and uniqueness of this equilibrium. Section 4 examines multi-period competition and derives the dynamic profit-neutrality result. Section 5 looks at the impact of the access charge on the social welfare. Section 6 discusses the implications of the model for policy. Section 7 concludes.

2 The model

There are two networks, 1 and 2. The two networks have full coverage and compete for a unit mass of consumers. The main elements are as follows:

Cost structure. The fixed cost to serve each subscriber is f, whereas c_O and c_T denote the marginal cost of providing a telephone call borne by the originating and terminating networks. The marginal cost of an on-net call is then $c \equiv c_O + c_T$. Network operators pay

[7] They also study competition with regulated or contractually determined reception charges and competition with endogenous reception charges and termination-based price discrimination. Their analysis shows that in the presence of termination-based price discrimination, direct externalities on the customers attached to the rival network create strong incentives for connectivity breakdown.

each other a reciprocal termination charge a when a call initiated on a network is terminated on a different network.[8] As a result, the access mark-up is equal to:

$$m \equiv a - c_T.$$

Retail pricing. I consider competition in nonlinear pricing. I do not, however, allow firms to charge different prices for on-net and off-net calls. Jeon et al. (2004) analyze this type of price discrimination and show that it creates strong incentives for connectivity breakdown (even among equal networks) in the presence of the receiver-pays regime. As the customers' demand function (for making and receiving calls) is known and the same for all customers that subscribe to a given network, firms cannot do better than offering three-part tariffs $\{F_i, p_i, r_i\}$, where F_i is the monthly subscriber charge, p_i is the per-unit calling price and r_i is the per-unit reception charge.

Market shares. The networks (*i.e.,* firms) sell a differentiated but substitutable product: customers are uniformly distributed on the segment $[0, 1]$, whereas the two networks are located at the two extremities of this segment, with 1 located at 0 and 2 located at 1. Consumers' tastes for networks are thus represented by their position on the segment and taken into account through the transportation cost t. Given income y, a consumer located at x and joining network i has utility

$$y + v_0 - t\,|x - x_i| + w_i,$$

where v_0 represents a fixed surplus from being connected to either network (it is assumed to be large enough so that all consumers want to subscribe to one network), $t\,|x - x_i|$ is the cost of subscribing to a network with "address" x_i, and w_i is the net surplus of a network i's subscriber from making and receiving calls on that network.

I am interested in allowing for asymmetric firms so that one firm may have a greater market share than the other even when their prices are the same. I am also interested in how access charges affect the firms' trade-off between charging low prices to build a locked-in customer base and charging high prices to exploit locked-in customers. Klemperer (1987) presents a model that considers these two facts and in which firms are differentiated à la Hotelling (as in the standard model of network competition).[9]

Therefore, as in Klemperer (1987) I assume that every customer incurs a cost $s \geq 0$

[8] Reciprocity means that a network pays as much for termination of a call on the rival network as it receives for completing a call originated on the rival network, i.e., access charges are symmetric or uniform.

[9] There is a large literature on switching costs, see Klemperer (2007, Chap. 2) for an overview of switching costs and competition.

when switching networks. Actually, there is much evidence suggesting that switching costs are significant (see for example Grzybowski and Pereira, 2007). I assume also that $s < t$, which implies that in a symmetric equilibrium some customers switch of network. For given net surpluses w_1 and w_2 offered by the two firms, market shares are determined as follows: at the very beginning there is a fraction α_0 of consumers attached to network 1. Of these, a consumer located at $x \in [0,1]$ will remain associated with network 1 if $w_1 - tx \geq w_2 - t(1-x) - s$. A consumer x attached to network 2 will instead switch to network 1 if $w_1 - tx - s \geq w_2 - t(1-x)$. Therefore, 1's market share is

$$\begin{aligned} \alpha_1 &= \alpha_0 \left[\frac{1}{2} + \sigma(w_1 - w_2 + s)\right] + (1-\alpha_0)\left[\frac{1}{2} + \sigma(w_1 - w_2 - s)\right] \\ &= \frac{1}{2} + (2\alpha_0 - 1)\sigma s + \sigma(w_1 - w_2), \end{aligned} \tag{1}$$

where $\sigma \equiv 1/2t$ measures the degree of substitutability between the two networks. As the two networks have full coverage, 2's market share is $\alpha_2 = 1 - \alpha_1$. The definition of market shares implicitly assumes that both "installed bases" will be shared. Nonetheless, for large enough price differentials, it can be the case that all consumers already attached to a given network stay with that network, whereas only a fraction of the consumers initially attached to the rival network switch of network. Then, (1) will be incorrect.[10] Indeed, (1) represents the 1's market share when and only when t is large enough (or equivalently σ is small enough), namely, when

$$|w_i - w_j| + s \leq t.^{11} \tag{2}$$

As pointed out by Klemperer (1987), a firm may find it profitable to deviate from its strategy in the candidate equilibrium defined by the first-order conditions by choosing a net surplus outside the range in which (2) holds. Nonetheless, the two networks will find no profitable to charge prices so that (2) fails to hold when they are *relatively* poor substitutes. The reason is that the firm's benefit of undercutting its rival depends on the degree of substitutability between the two networks: $dx/dw_1 = \sigma$ (consider for instance the extreme case where σ is close to zero). Thus, for a large enough t the two networks will share the customer base. As I will show in Section 3, the assumption $s < t$ ensures that firms have no incentive to charge prices outside the range in which (2) holds. Finally, notice that either $\alpha_0 = 1/2$ or $s = 0$

[10] For example, in the first line of equation (1), the term in the first bracket could exceed 1, in which case it should be "truncated", even if the term in the second bracket lies between 0 and 1.

[11] Notice that $1/2 + \sigma(w_i - w_j + s) < 1$ if $w_i + s - w_j < t$, whereas $1/2 + \sigma(w_i - w_j + s) > 0$ if $w_i - w_j + s > -t$, or, equivalently, if $w_j - w_i - s < t$. These two conditions simplify to $|w_i - w_j + s| < t$. Similarly, $1/2 + \sigma(w_i - w_j - s) < 1$ if $w_i - w_j - s < t$, whereas $1/2 + \sigma(w_i - w_j - s) > 0$ if $w_i - w_j - s > -t$ or, equivalently, if $w_j + s - w_i < t$. These two conditions simplify to $|w_j - w_i + s| < t$.

will yield symmetry between the networks.

Individual demand. Subscribers obtain positive utility from making and receiving calls. The caller's utility in making a call of length q minutes is $\mu(q)$, whereas the receiver's is $\widetilde{\mu}(q)$ for receiving a call of that length. I assume that these utility functions $\mu(\cdot)$ and $\widetilde{\mu}(\cdot)$ are twice continuously differentiable, with $\mu' > 0$, $\mu'' < 0$, $\widetilde{\mu}' > 0$ and $\widetilde{\mu}'' < 0$, which implies that the customers' demand function is differentiable (I assume, as do Jeon et al. 2004, that the marginal surplus from receiving a call is nonnegative).

The caller's demand function $q(\cdot)$ is given by $\mu'(q) = p$, whereas the receiver's demand function $\widetilde{q}(\cdot)$ is given by $\widetilde{\mu}'(\widetilde{q}) = r$. When receivers are allowed to hang up, the length of a call from a caller of network i to a receiver of network j is given by $Q(p_i, r_j) = \min\{q(p_i), \widetilde{q}(r_j)\}$. The model is then discontinuous and complicates the analysis.[12] In order to get around this problem, I assume that both the caller's and receiver's utilities are subject to a random noise, which smooths the demand. Let ε and $\widetilde{\varepsilon}$ denote the random term of the caller's and receiver's utilities. I assume that: i) ε and $\widetilde{\varepsilon}$ respectively follow the distribution functions $F(\cdot)$ and $\widetilde{F}(\cdot)$, with supports $[\underline{\varepsilon}, \overline{\varepsilon}]$ and $[\underline{\widetilde{\varepsilon}}, \overline{\widetilde{\varepsilon}}]$, where $\overline{\varepsilon} - \underline{\varepsilon} > 0$ and $\overline{\widetilde{\varepsilon}} - \underline{\widetilde{\varepsilon}} > 0$, and strictly positive density functions $f(\cdot)$ and $\widetilde{f}(\cdot)$; ii) these random terms are identically and independently distributed for each caller-receiver pair. Given this, I make the following assumption:

A.1. The caller's and receiver's utilities are linear functions of the random term: $u = \mu(q) + \varepsilon q$ and $\widetilde{u} = \widetilde{\mu}(\widetilde{q}) + \widetilde{\varepsilon}\widetilde{q}$.

Assumption A.1. allows the willingness to stay on the phone to be state-contingent for both callers and receivers. I further assume that the demands q and $\widetilde{q}$ are bounded.[13] As a result, the profit functions are also bounded ($\alpha_i \in [0, 1]$).

Under A.1, and for a given pair of prices (p_i, r_j), the average length of a call from a network i's subscriber to a network j's subscriber is:

$$Q(p_i, r_j) = \int_{\underline{\varepsilon}}^{\overline{\varepsilon}} \int_{\underline{\widetilde{\varepsilon}}}^{\overline{\widetilde{\varepsilon}}} [q(p_i, \varepsilon)\boldsymbol{\Upsilon}_{ij} + \widetilde{q}(r_j, \widetilde{\varepsilon})(1 - \boldsymbol{\Upsilon}_{ij})] f(\varepsilon)\widetilde{f}(\widetilde{\varepsilon}) d\varepsilon d\widetilde{\varepsilon},$$

where Υ_{ij} is an indicator variable taking the value 1 if $q(p_i, \varepsilon) \leq \widetilde{q}(r_j, \widetilde{\varepsilon})$ and 0 otherwise. The

[12] When reception charges are regulated or contractually determined before the firms compete in retail tariffs, an assumption that simplifies much the analysis is that the caller determines the volume of calls. However, when reception charges are endogenous, this assumption introduces a potential problem in the analysis due to the *multiplicity of equilibria:* from the viewpoint of networks and subscribers, only the sum $\{F_i + r_i\widetilde{q}\}$ matters, not its composition. As a result, different combinations of F_i and r_i are feasible equilibria but nonequivalent since each combination may affect differently the rival network.

[13] For a given $\varepsilon \in [\underline{\varepsilon}, \overline{\varepsilon}]$, there exist calling charges $\underline{p}$ and $\overline{p}$ so that $q = \overline{q}$, with $0 < \overline{q} < \infty$, for $p \leq \underline{p}$, and $q = 0$ for $p \geq \overline{p}$. Similarly, for a given $\widetilde{\varepsilon} \in [\underline{\widetilde{\varepsilon}}, \overline{\widetilde{\varepsilon}}]$ there exist reception charges $\underline{r}$ and $\overline{r}$ so that $\widetilde{q} = \overline{\widetilde{q}}$, with $0 < \overline{\widetilde{q}} < \infty$, for $r \leq \underline{r}$, and $\widetilde{q} = 0$ for $r \geq \overline{r}$.

average utility that a network i's subscriber obtains from calling a network j's subscriber is then

$$U(p_i, r_j) = \int_{\underline{\varepsilon}}^{\overline{\varepsilon}} \int_{\underline{\widetilde{\varepsilon}}}^{\overline{\widetilde{\varepsilon}}} [u(q(p_i, \varepsilon))\mathbf{\Upsilon}_{ij} + u(\widetilde{q}(r_j, \widetilde{\varepsilon}))(\mathbf{1} - \mathbf{\Upsilon}_{ij})] f(\varepsilon)\widetilde{f}(\widetilde{\varepsilon}) d\varepsilon d\widetilde{\varepsilon},$$

whereas the average utility that a network j's subscriber obtains from receiving a call from a network i's subscriber is

$$\widetilde{U}(p_i, r_j) = \int_{\underline{\varepsilon}}^{\overline{\varepsilon}} \int_{\underline{\widetilde{\varepsilon}}}^{\overline{\widetilde{\varepsilon}}} [\widetilde{u}(q(p_i, \varepsilon))\mathbf{\Upsilon}_{ij} + \widetilde{u}(\widetilde{q}(r_j, \widetilde{\varepsilon}))(\mathbf{1} - \mathbf{\Upsilon}_{ij})] f(\varepsilon)\widetilde{f}(\widetilde{\varepsilon}) d\varepsilon d\widetilde{\varepsilon}.$$

Therefore, the average volume of traffic from network i to network j depends on the two usage prices, and the length of a call is sometimes determined by the caller and at other times by the receiver. In this general setup, the following useful results hold:

$$\frac{\partial Q(p_i, r_j)}{\partial p_i} = \int_{\underline{\widetilde{\varepsilon}}}^{\overline{\widetilde{\varepsilon}}} \int_{\underline{\varepsilon}}^{\overline{\varepsilon}} \left(\frac{\partial q(p_i - \varepsilon)}{\partial p_i} \mathbf{\Upsilon}_{ij} \right) f(\varepsilon)\widetilde{f}(\widetilde{\varepsilon}) d\varepsilon d\widetilde{\varepsilon},$$

$$\frac{\partial U(p_i, r_j)}{\partial p_i} = p_i \frac{\partial Q(p_i, r_j)}{\partial p_i}, \tag{3}$$

where I have used $\partial u(q(p_i - \varepsilon))/\partial q = \mu'(\mu'^{-1}(p_i - \varepsilon)) + \varepsilon = p_i$. Similarly,

$$\frac{\partial Q(p_j, r_i)}{\partial r_i} = \int_{\underline{\widetilde{\varepsilon}}}^{\overline{\widetilde{\varepsilon}}} \int_{\underline{\varepsilon}}^{\overline{\varepsilon}} \left(\frac{\partial \widetilde{q}(r_i - \widetilde{\varepsilon})}{\partial r_i} \mathbf{\Upsilon}_{ji} \right) f(\varepsilon)\widetilde{f}(\widetilde{\varepsilon}) d\widetilde{\varepsilon} d\varepsilon,$$

$$\frac{\partial \widetilde{U}(p_j, r_i)}{\partial r_i} = r_i \frac{\partial Q(p_j, r_i)}{\partial r_i}, \tag{4}$$

where I have used $\partial \widetilde{u}(\widetilde{q}(r_i - \widetilde{\varepsilon}))/\partial \widetilde{q} = \widetilde{\mu}'(\widetilde{\mu}'^{-1}(r_i - \widetilde{\varepsilon})) + \widetilde{\varepsilon} = r_i$. For the sake of the presentation, I denote $Q_{ij} \equiv Q(p_i, r_j)$, $U_{ij} \equiv U(p_i, r_j)$ and $\widetilde{U}_{ij} \equiv \widetilde{U}(p_i, r_j)$ $\forall i, j$. I make the standard assumption of a balanced calling pattern,[14] which means that the percentage of calls originating on a network and completed on the same network is equal to the fraction of consumers subscribing to the network. The net surplus is then given by

$$w_i = \phi_i - F_i, \tag{5}$$

[14] Dessein (2003, 2004) examines the impact of unbalanced calling patterns between different customer types on retail competition when network operators compete in the presence of the caller-pays regime.

with

$$\phi_i(\alpha_i, p_i, p_j, r_i, r_j) = \alpha_i U_{ii} + \alpha_j U_{ij} + \alpha_i \widetilde{U}_{ii} + \alpha_j \widetilde{U}_{ji} \\ -p_i(\alpha_i Q_{ii} + \alpha_j Q_{ij}) - r_i(\alpha_i Q_{ii} + \alpha_j Q_{ji}). \quad (6)$$

Timing. The timing of the game is the following: 1. The reciprocal access charge is set by a regulator or negotiated between the two networks; 2. The two networks simultaneously choose retail prices; 3. Consumers make subscription and consumption decisions.

3 Price competition

In this section, I consider competition in the presence of the receiver-pays regime and for a given reciprocal access charge. The profit of network i can be written as (for $i \neq j = 1, 2$):

$$\pi_i = \alpha_i[\alpha_i(p_i - c)Q_{ii} + \alpha_j(p_i - c - m)Q_{ij} + \alpha_j m Q_{ji} \\ +r_i(\alpha_i Q_{ii} + \alpha_j Q_{ji}) + F_i - f]. \quad (7)$$

The network i maximizes π_i with respect to p_i, r_i and F_i. We can solve the firm's problem by maximizing π_i with respect to p_i and r_i for a given α_i, while adapting F_i so as to maintain net surpluses w_1 and w_2 and therefore market shares constant. For this to hold, the net surpluses must satisfy $w_i - w_j = (1/\sigma)(\alpha_i - 1/2) - (2\lambda_i - 1)s$, where $\lambda_1 = \alpha_0$ and $\lambda_2 = 1 - \alpha_0$; using (5) we have

$$F_i = \phi_i - \phi_j + F_j - \frac{1}{\sigma}\left(\alpha_i - \frac{1}{2}\right) + (2\lambda_i - 1)s.$$

After substitution of equation F_i into the profit function, we have

$$\pi_i(p_i, r_i, \alpha_i) = \alpha_i[\alpha_i(p_i - c)Q_{ii} + \alpha_j(p_i - c - m)Q_{ij} + \alpha_j m Q_{ji} + r_i(\alpha_i Q_{ii} + \alpha_j Q_{ji}) \\ +\phi_i - \phi_j + F_j - \frac{1}{\sigma}(\alpha_i - \frac{1}{2}) + (2\lambda_i - 1)s - f]. \quad (8)$$

Characterization of the off-net-cost pricing equilibrium

For given $r_i = r_j = r$ and p_j, the calling charge p_i determines the volume of calls when network i's callers are sovereign. Network i incurs a unit cost $c + \alpha_j m$ from delivering these calls to network i and network j. Also, p_i affects the subscribers' net surplus, so fixed fees must be adapted in order to maintain markets shares constant. More precisely, a decrease in p_i:

- affects network i's revenue, but at the expense of consumers. Thus, fixed fees must be adapted so as to neutralize this transfer and keep market shares constant.
- allows network i to increase its fixed fee by $U(p_i, r)$, which is the utility that network i's subscribers obtain from making calls (the calling charge also affects the utility from incoming calls, but fixed fees do not need to be adapted to maintain market shares because i's and j's subscribers are affected in the same way.)
- affects the amount of money that network j's subscribers pay for the calls received from network i. This effect is called *pecuniary externality* in Jeon et al. (2004), and allows network i to increase its fixed fee by $r\alpha_i$, while keeping market shares constant.

Therefore, network i sets p_i so as to maximize:

$$[r\alpha_i - (c + \alpha_j m)]Q(p_i, r) + U(p_i, r). \tag{9}$$

For given $p_i = p_j = p$ and r_j, the reception charge r_i determines the volume of calls when network i's receivers are sovereign. For this volume of calls, network i incurs a cost $\alpha_i c$, but gains $\alpha_j m$ from the off-net calls. The reception charge also affects subscribers' net surpluses:

- network i gains revenue from reception charges, its fixed fee must then be altered by the same amount to keep market shares constant.
- network i can increase its fixed fee to reflect the utility from receiving calls: $\widetilde{U}(p, r_i)$ (reception charges affect similarly network i's and j's subscribers for the calls they place on network i: $\alpha_i U(p, r_i)$, so fixed fees must not be adapted.)
- a *pecuniary externality:* the reception charge r_i affects the amount of money that network j's subscribers pay for calling network i's subscribers. This externality allows network i to increase its fixed fee by $p\alpha_i$ while keeping market shares constant.

Therefore, network i sets r_i so as to maximize:

$$[\alpha_j m + p\alpha_i - \alpha_i c]Q(p, r_i) + \widetilde{U}(p, r_i). \tag{10}$$

By differentiating (9) with respect to p_i and (10) with respect to r_i, and using (3) and (4) we obtain the following first-order conditions:

$$p_i = c + \alpha_j m - \alpha_i r, \tag{11}$$

$$r_i = \alpha_i c - \alpha_j m - \alpha_i p. \tag{12}$$

Using the terminology of Jeon et al. (2004), we have that network operators charge calls and call receptions at the *strategic marginal cost:* the i's equilibrium calling price is equal to the average unit cost of a call originating on network i minus the pecuniary externality imposed on network j's subscribers; likewise, the network i's equilibrium reception charge is equal to the average unit cost of receiving calls on network i minus the pecuniary externality imposed on network j's subscribers. If $p_i = p$ and $r_i = r$, equations (11) and (12) simplify to

$$p = c + m, \tag{13}$$

$$r = -m. \tag{14}$$

That is, network operators charge calls and call receptions at their off-net cost. I shall emphasize that this symmetric solution is valid for any given level of market shares. Therefore, if in the beginning firms are asymmetric (in terms of market shares), they still will charge calls and call receptions at their off-net cost. This is the so-called 'off-net-cost pricing principle': each network sets prices for a subscriber's outgoing and incoming traffic at the marginal cost that it would incur if *all* other subscribers belonged to the rival network. The off-net-cost pricing principle dates back to Laffont et al. (2003), who found this pricing rule in a framework for Internet backbone competition. In contrast, Jeon et al. (2004) and this article analyze three-part tariff competition in a telecommunications environment. My framework however generalizes their work by allowing a random noise in both the callers and receivers' utilities, and by removing the assumption of a given proportionality between the utility functions.

The profit-neutrality result

I now show that in the presence of the receiver-pays regime, equilibrium profits are independent of the access charge. Equations (13) and (14) characterize the equilibrium usage prices, which are symmetric. By setting calling and reception charges at the off-net cost (Proposition 1 provide sufficient conditions for the existence and uniqueness of this equilibrium), i's profit can be rewritten as follows (for $i \neq j = 1, 2$):

$$\pi_i = \left(\frac{1}{2} + (2\lambda_i - 1)\sigma s - \sigma(F_i - F_j)\right)(F_i - f). \tag{15}$$

The first-order condition yields

$$F_i = \frac{1}{2}\left[f + \frac{1}{2\sigma} + (2\lambda_i - 1)s + F_j\right]. \tag{16}$$

Thus, the equilibrium fixed fees are

$$F_i = f + \frac{1}{2\sigma} + (2\lambda_i - 1)\frac{s}{3}. \tag{17}$$

Substituting F_1 and F_2 into (15) yields the equilibrium profit

$$\pi_i = \frac{1}{4\sigma} + (2\lambda_i - 1)\frac{s}{3}\left[1 + (2\lambda_i - 1)\frac{\sigma s}{3}\right]. \tag{18}$$

Static Profit-Neutrality Result. Equation (18) shows that at a symmetric equilibrium with symmetric customer bases the profit is equal to the profit that each network would obtain under unit demands, i.e., $1/4\sigma$. Also, we have that the equilibrium profit does not depend on the level of the access charge. Intuitively, consider a given access markup m. The off-net calls are charged at their marginal cost, so the profit from these calls is always zero independently of the level of the access markup. The price of an on-net call is $p = c + m$, whereas its marginal cost is c; thus, the profit from these calls is $m\alpha_i Q$. In addition, network i gains $m\alpha_j Q$ in the wholesale market. Therefore, the total gain from placing calls and terminating off-net calls is mQ. Consider now receptions, network i charges $r = -m$, so it obtains $-m\alpha_i Q$ from the on-net calls and $-m\alpha_j Q$ from the off-net calls. Therefore, the i's profit from receiving calls is $-mQ$. This gain cancels out the profit from placing calls and terminating rivals' calls. As a result, the equilibrium profit is independent of the level of the access charge. Laffont et al. (1998a) obtain a similar (static) profit-neutrality result but in a different framework: the two networks compete in two-part tariffs and in the absence of reception charges. The intuition of their profit-neutrality result is also different: consider an increase in the access charge, it raises the usage price, which in turn makes it more desirable for networks to build market share. In the linear pricing case, networks cannot build market share without incurring an access deficit,[15] however under two-part tariffs they can do it by lowering their fixed fees while keeping usage prices constant. Actually, networks find it worthwhile to spend the full termination revenue to build market share, so that the access charge no longer affects equilibrium profits. Gans and King (2001) show that in the presence of termination-based price discrimination the waterbed effect is higher than one hundred percent so that firms favour a below-cost access charge.

[15] As a result, firms can use the access charge to increase profits.

Existence and uniqueness of the off-net-cost pricing equilibrium

To prove the existence and uniqueness of the off-net-cost pricing equilibrium, I have to be more specific about the distribution function of the noise and the caller's and receiver's utility functions (this does not mean that the off-net-cost pricing equilibrium does not exist for other distribution or utility functions). I will make the following assumption:

A.2. $\mu(q) = aq - (b/2)q^2$ and $\widetilde{\mu}(\widetilde{q}) = d\widetilde{q} - (e/2)\widetilde{q}^2$, where $a, b, d, e > 0$. Moreover, $\varepsilon, \widetilde{\varepsilon} \in [\underline{\varepsilon}, \overline{\varepsilon}]$, where $\underline{\varepsilon} < 0 < \overline{\varepsilon}$, $E(\varepsilon) = E(\widetilde{\varepsilon}) = 0$, and both random terms follow a uniform distribution with density function: $f(\varepsilon) = \widetilde{f}(\widetilde{\varepsilon}) = 1/\Delta$, where $\Delta = \overline{\varepsilon} - \underline{\varepsilon} > 0$.

Notice that A.1. and A.2. imply linear demand functions: $q = (a - p + \varepsilon)/b$ and $\widetilde{q} = (d - r + \widetilde{\varepsilon})/e$.

Proposition 1 *(Existence and Uniqueness) Under A.1. and A.2., for a small enough σ and a large enough Δ there exists a unique equilibrium, which is interior and where the network $i = 1, 2$ charges $p_i = c + m$, $r_i = -m$ and $F_i = f + t + (2\lambda_i - 1)s/3$, with $\lambda_1 = \alpha_0$ and $\lambda_2 = 1 - \alpha_0$.*

Proof. See Appendix. ■

As commented above, Jeon et al. (2004) also analyze competition in the presence of the receiver-pays regime and establish the existence of the off-net-cost pricing equilibrium in the specific case where the noise on the receiver side converges to zero, so that the length of a call is determined by the caller with probability converging to one. Instead, Proposition 1 says that the off-net-cost pricing equilibrium exists and is unique for a small enough σ and a large enough Δ. Indeed, a small (enough) σ, which means that networks are *relatively* poor substitutes, is a standard assumption in the "two-way" access literature. As Laffont et al. (1998a) pointed out, it is not a mere technical problem but a robust economic problem, namely when the two networks are close enough substitutes (and switching costs are not too high), each network has an incentive to undercut its rival to corner the market. A cornered-market outcome cannot be an equilibrium however. The reason is that if one network makes a positive profit, the other could mimic it and obtain positive profit provided that $s < t$ (see proof of Proposition 1). A small (enough) σ does not seem to be a too restrictive assumption. For example, products are differentiated by quality of voice service (grade of service and quality of service), brand image and customer service. A large (enough) Δ does not seem to be a too restrictive assumption either as extreme situations might occur in reality: there exist many situations in which a person may not want to receive a call even though it is free or, conversely, has to send/receive a message even though it is too expensive.

The conditions provided in Proposition 1 ensure that each firm's profit function is concave in its own strategies in the region in which (1) holds. But this is not sufficient to show that the solution to the first-order conditions defines an equilibrium because a firm may find it profitable to deviate from this candidate equilibrium by choosing a strategy outside the range given by equation (2). Nonetheless, this deviation is not profitable. Consider the candidate equilibrium characterized by (13), (14) and (17), and suppose that network $i \neq j = 1, 2$ deviates so as to corner the λ_i segment of the market:[16] $1/2 + \sigma(F_j - F_i + s) > 1$, which requires that

$$F_i < F_j - t + s, \tag{19}$$

or the λ_j segment of the market: $1/2 + \sigma(F_j - F_i - s) > 1$, which requires that

$$F_i < F_j - t - s, \tag{20}$$

where $F_j = f + t + (2\lambda_j - 1)s/3$. We can consider only the equation (19), i.e., $F_i < f + (2\lambda_j - 1)s/3 + s$, because it is less strict than equation (20). In other words, if (19) fails to hold, then (20) will also fail to hold. For the deviation to be profitable, i must not gain from a marginal increase in its fee, which amounts to:

$$0 \geq \left.\frac{\partial \pi_i}{\partial F_i}\right|_{F_1, F_2 : \alpha_i = 1} = -\sigma(F_i - f) + 1,$$

that is, $F_i \geq f + 2t$. Therefore, i's fee must lie in the range

$$f + 2t < F_i < f + (2\lambda_j - 1)\frac{s}{3} + s. \tag{21}$$

If $\lambda_j = 0$, (21) reduces to $s > 3t$, which is not feasible since $s < t$. Instead, if $\lambda_j = 1$, (21) reduces to $s > 3t/2$, which is not feasible either.

4 Multi-period competition

To provide a motivation for the analysis of the multi-period model, it is convenient to introduce briefly the main insights of a companion work, López (2007a), that extends the standard model of static competition (in which the caller-pays regime is assumed throughout, and networks compete in two-part tariffs and in the absence of termination-based price

[16] We can consider only deviations in fixed fees from the shared-market candidate equilibrium because firms find it optimal to charge calls and call receptions at their off-net cost independently of the level of the market shares, furthermore the off-net prices are the unique solution to the first-order conditions when the conditions provided in Proposition 1 are satisfied.

discrimination), to a two-period model. López (2007a) shows that even symmetric networks with consumers' full participation[17] can use reciprocal access charges to undermine competition, so the profit-neutrality result of Laffont et al. (1998a) no longer holds. I briefly provide the intuition of this result but refer to López (2007a) for a detailed discussion. Consider the static model but with asymmetric networks, Carter and Wright (2003) show that the profit of the larger network decreases when the access charge departs away from the marginal cost.[18] Consider now symmetric networks and a two-period model. In the second period the model is similar to the static model, thus the profit of being large in the second-period decreases when the access charge departs away from the marginal cost, which in turn lowers the incentive to fight for market share in the first period. This result restores the idea that a collusion concern could be associated with an above-cost access charge.[19]

The question that I address in this section is wether the 'dynamics' could also break the profit-neutrality result that competition in the presence of reception charges yields. Conversely, I now show that competition in calling prices, reception charges and fixed fees neutralizes the impact of the access charge on multi-period competition.

To simplify the analysis, I make the following two assumptions:

A.3. *Preferences are independent across periods.*

A.4. *Consumers have naive expectations.*

A.3. says that preferences may change over time. Alternatively, and following Klemperer (1987), I can consider the more general case where a fraction $v < 1$ of second-period consumers are new in the market, and a fraction $\mu > 0$ and $1 - \mu - v$ of first-period consumers have, respectively, independent and unchanged preferences across periods. This more general framework only changes the definition of the market shares because it only affects how the current market share depends on the installed customer bases. Nonetheless, the game can be solved by maximizing profits with respect to usage prices, while adapting fixed fees so that market shares remain constant, which implies that the equilibrium usage prices are not affected by these more general preferences. Put another way, the off-net-cost pricing result is independent of the market shares, so it cannot be affected by their definition. Therefore, the

[17] Dessein (2003) shows that if the subscription demand is elastic, firms prefer an access charge below marginal cost.

[18] The reason is that the large network faces a higher proportion of on-net calls, whereas the small network faces a higher proportion of off-net calls. Then, an above-cost access charge raises the average unit cost of the small network. Networks charge calls at the average unit cost, thus it follows that the large network will face a net outflow of calls and as a result a deficit in the wholesale market. Conversely, if the access charge is below marginal cost, the large network will face a net inflow of calls because the small network will face a lower average unit cost. A net inflow of calls alongside a below-cost access charge is not profitable.

[19] Actually, also in the presence of a below-cost access charge, however this practice is limited by feasibility constraints: the access charge cannot be negative, the lower bound of m is then $-c_T$.

main results of this article, namely the (static and dynamic) profit-neutrality results, will also hold in this more general setup. The equilibrium fixed fees may be different to the ones obtained in the case where all the customers have independent preferences across periods, however this does not affect the profit-neutrality results. A.4. imposes a strong condition on the consumer behaviour. Nonetheless, as I will argue later, assuming rational consumer expectations would not affect the main insights either.

Consider a two-period model. Networks have rational expectations and discount future revenues and costs by a factor δ, in the first-period i solves:[20]

$$\max_{p_i^1, w_i^1} \Pi \equiv \pi_i^1(p_i^1, p_j^1, w_i^1, w_j^1) + \delta \pi_i^2(\alpha_0(w_i^1, w_j^1)),$$

where π_i^1 is the first-period profit and π_i^2 is the equilibrium second-period profit of network i, which is given by (18) and is a function of the customer base α_0 but not of the access mark-up m. The equilibrium first-period prices are (for $i = 1, 2$):

$$p_i^1 = c + m, \quad r_i^1 = -m, \tag{22}$$

and

$$F_i^1 = f + \frac{1}{2\sigma} - \frac{2s\delta}{3}. \tag{23}$$

Thus,

$$\Pi = \frac{1+\delta}{4\sigma} - \frac{s\delta}{3}.$$

Dynamic Profit-Neutrality Result. López (2007a) shows that when network operators compete in calling prices and fixed fees: i) the expressions $\partial \pi_1^2 / \partial \alpha_0$ and $\partial \pi_2^2 / \partial (1 - \alpha_0)$ depend on both α_0 and m, and ii) at a symmetric equilibrium, slightly moving m away from zero lowers the value of having a higher market share in the second period, $\partial \pi_1^2 / \partial \alpha_0$ and $\partial \pi_2^2 / \partial (1 - \alpha_0)$ are strictly concave in m at $m = 0$, which in turn softens first-period competition and therefore raises the networks' total discounted profit. In contrast, in the presence of reception charges the expression π_i^2 does not depend on the second-period access charge independently of the sizes of the installed bases (the intuition behind this last result is given in the previous section.) As a result, we have that $\partial^2 \pi_i^2 / \partial m \partial \alpha_0 = 0$ for any α_0, so the first-period fixed fees and the total discounted profits are independent of the level of the access charge.

Existence and Uniqueness. The discussion paper López (2007b) shows that for a small

[20] As commented above, it is analytically convenient to view network competition as one in which the networks pick calling prices and net surpluses rather than calling prices and fixed fees, so that market shares are determined directly by net surpluses.

enough σ and a large enough Δ the unique interior equilibrium is characterized by (22) and (23), and that there exists no "cornered-market" equilibrium provided that switching costs are small enough.

Rational Consumer Expectations. In the second period, the equilibrium prices are the same for both naive and rational consumer expectations: (13), (14) and (17). In the first period, however, consumers with rational expectations recognize that a network with higher market share will charge higher prices in the second-period provided that switching costs exist. This consumers' realization makes demand less elastic and thus softens first-period competition. Consequently, first-period fixed fees are higher in the presence of rational consumer expectations than in their absence. Nevertheless, as the value of having a higher second-period market share is independent of the access charge, first-period prices do not depend on the access charge either, and hence the access charge has no impact on the subscribers' first-period net surpluses. Therefore, when firms compete for market share in fixed fees and calling and reception charges, their total discounted profit is not affected by the access charge in the presence of both naive and rational consumers expectations.

A Multi-Period Model. This two-period model can be easily extended to a model in which firms compete in (finite) T discrete periods of time. López (2007b) shows that in such a multi-period model, for a small enough σ and a large enough Δ there exists an interior subgame-perfect equilibrium so that in any continuation equilibria (even off the equilibrium path): (i) networks charge calls at their off-net cost, (ii) the fixed fees and per-period profits depend on the history of play until that date only through the installed customer bases, and do not depend on the termination mark-up levels.

5 Social Optimum

Jeon et al. (2004) already pointed out that efficiency cannot be achieved in the presence of a (nonvanishing) noise. The reason is that the price instruments must be contingent on the realization of this random term because it affects the marginal utilities. I address this problem in a different way, I look for the level of the access charge that maximizes the *average* social welfare.

As there is full participation and payments are only transfers from one agent to another, from a social welfare perspective, what matters is the utility that consumers derive from incoming and outgoing calls, and the cost of these calls. Consider a call from a network i's subscriber to a network j's subscriber, the average length of this call is Q_{ij}, whereas the average utility derived from this call is $U_{ij} + \widetilde{U}_{ij}$. Let W denote the average social welfare of such a call, which is a function of the access mark-up. In equilibrium

$$W(m) = U(c+m,-m) + \widetilde{U}(c+m,-m) - cQ(c+m,-m). \tag{24}$$

Notice that equation (24) also reflects the average total utility.[21] The first-order derivative is

$$\frac{dW}{dm}(m) = \left(\frac{\partial U}{\partial p}(m) - \frac{\partial U}{\partial r}(m)\right) + \left(\frac{\partial \widetilde{U}}{\partial p}(m) - \frac{\partial \widetilde{U}}{\partial r}(m)\right) - c\left(\frac{\partial Q}{\partial p}(m) - \frac{\partial Q}{\partial r}(m)\right). \tag{25}$$

For linear demands $q(p)$ and $\widetilde{q}(r)$, a low enough σ and a large enough Δ we have

$$d^2W/(dm)^2 \simeq -(1/2)(1/b + 1/e + e/b^2 + b/e^2) < 0. \tag{26}$$

Let m^* denote the socially optimal termination mark-up so that W is maximized. If (26) holds, then m^* solves $(dW/dm)(m^*) = 0$. A small increase in the access charge has two opposite effects: it raises the calling charge and it lowers the reception charge. A higher calling charge lowers the callers' willingness to stay on the phone, which in turn lowers both the callers' utility $\partial U/\partial p$ and the receivers' utility $\partial \widetilde{U}/\partial p$. Instead, a small decrease in the reception charge raises the receivers' willingness to stay on the phone, which in turn raises the utility of both callers and receivers: $-(\partial U/\partial r + \partial \widetilde{U}/\partial r)$. Therefore, on the one hand, the average volume of traffic in which callers are sovereign decreases, which lowers the cost by: $-c(\partial Q/\partial p)$. But on the other hand the average volume of traffic in which receivers are sovereign increases, which raises the cost by: $c(\partial Q/\partial r)$. Using (3) and (4) yields, in equilibrium, $\partial U/\partial p = (c+m)(\partial Q/\partial p)$ and $\partial \widetilde{U}/\partial r = -m(\partial Q/\partial r)$. Then, equation (25) simplifies to:

$$\frac{dW}{dm}(m) = m\left(\frac{\partial Q}{\partial p}(m) + \frac{\partial Q}{\partial r}(m)\right) + c\frac{\partial Q}{\partial r}(m) + \left(\frac{\partial \widetilde{U}}{\partial p}(m) - \frac{\partial U}{\partial r}(m)\right). \tag{27}$$

The value of the socially optimal termination mark-up depends on the characteristics of the market. Starting from $m = 0$, consider a small increase in the termination mark-up,

[21] The average utility that a network i's subscriber derives from making and receiving calls is $\alpha_i U_{ii} + \alpha_j U_{ij} + \alpha_i \widetilde{U}_{ii} + \alpha_j \widetilde{U}_{ji}$, whereas the cost incurred is $(\alpha_i Q_{ii} + \alpha_j Q_{ij})c$. Since there are α_i customers attached to network i and α_j customers attached to network j, the average total utility is

$$\alpha_i(\alpha_{ii}U_{ii}+\alpha_j U_{ij}+\alpha_i\widetilde{U}_{ii}+\alpha_j\widetilde{U}_{ji})+\alpha_j(\alpha_j U_{jj}+\alpha_i U_{ji}+\alpha_j\widetilde{U}_{jj}+\alpha_i\widetilde{U}_{ij})-\alpha_i(\alpha_i Q_{ii}+\alpha_j Q_{ij})c-\alpha_j(\alpha_j Q_{jj}+\alpha_i Q_{ji})c.$$

In equilibrium, this expression simplifies to (24) since $U_{ii} = U_{ij}$, $\widetilde{U}_{ij} = \widetilde{U}_{ji}$ and $Q_{ij} = Q_{ji}$. In words, the utility that customers derive from making and receiving calls is the same in the two networks even if they are asymmetric.

it slightly raises the calling charge and slightly lowers the reception charge; the first-order condition boils down to

$$\frac{dW}{dm}(0) = \left(c\frac{\partial Q}{\partial r}(0) + \frac{\partial \widetilde{U}}{\partial p}(0) \right) - \frac{\partial U}{\partial r}(0). \tag{28}$$

In general, (28) is different from zero.[22] Roughly speaking, $m = 0$ usually will not maximize the average social welfare of a given call. In particular, if $(dW/dm)_{m=0}$ is negative (positive) it follows that m^* is negative (positive) as long as the second-order condition (26) holds. Therefore, for a given call and cost-oriented access charges the sign of dW/dm depends on the impact of the calling price on the receivers surplus $\partial\widetilde{U}/\partial p$, and on the impact of the reception charge on the callers surplus $\partial U/\partial r$ and the length of the call $\partial Q/\partial r$.

If callers and receivers have the same utility function, i.e., $a = d$ and $b = e$, for a small enough σ and a large enough Δ we have that $(\partial\widetilde{U}/\partial p)_{m=0} - (\partial U/\partial r)_{m=0} \simeq -c/2b < 0$,[23] moreover (26) holds. Therefore, the consumers' surplus is locally decreasing in m: the lower reception charge raises the callers' utility less than the decrease in the receivers' utility due to the higher calling charge. This social cost together with the industry cost yield $(dW/dm)_{m=0} \simeq -c/b < 0$. Conversely, a small decrease in m lowers p and raises r so that the consumers' surplus increase. Moreover, a higher reception charge, lowers the industry cost by $|c\partial Q/\partial r|$. Therefore, W increases as m decreases below zero. We thus have

Proposition 2 *If callers and receivers have the same utility functions, then the access charge that maximizes the average social welfare of a given call W is below cost and given by (27) as long as $a^* = m^* + c_T > 0$. Otherwise, bill and keep, i.e., $a = 0$, maximizes W and as a result, in equilibrium the calling charge reflects the cost of originating a call $p = c_0$, whereas the reception charge reflects the cost of terminating a call $r = c_T$.*

So far, the consumers' disutility from not being able to join to their preferred network and the social costs of switching of network have been omitted from the analysis. If installed customer bases are unequal and every subscriber incurs a cost when switching of network, then in equilibrium symmetric market shares are not necessarily optimal. The equilibrium net surpluses, which determine market shares, depend only on the customer bases but not on the termination mark-up. Consequently, in order to obtain a socially optimal configuration of market shares, we would need an additional instrument or a direct regulation of the fixed fees.

[22] For a small enough σ and a large enough Δ, $\partial\widetilde{U}/\partial p$ and $\partial U/\partial r$ can be rewritten as functions of the parameters a, b, d, e, the marginal cost c and the access mark-up m (see Proposition 3 of the discussion paper López, 2007b).

[23] For more details see Proposition 3 of the discussion paper López (2007b).

6 Discussion and policy implications

The analysis shows that network operators charge incoming calls only when access charges are below cost, this result is consistent with the U.S. experience. Likewise, the analysis shows that network operators will pay subscribers for receiving calls (i.e., charge negative reception charges) when access charges are above cost. This is not common practice, however this is not to say that the model is incorrect. As long as the volume of calls is jointly determined by the caller and the receiver, and access charges are above cost, firms may find it optimal to pay customers for receiving calls. For example, in 2006 the new UK mobile operator H3G announced that it will pay its subscribers 5 ppm for receiving calls; Armstrong and Wright (2008) remark that this offer was surely motivated by its high termination charge.[24] Several reasons may explain why firms are usually reluctant to pay customers for receiving calls when it is optimal to do so: first, in order to avoid a price war that could dissipate the termination revenues (as my model predicts the firms' profit is independent of the access charge when network operators charge or pay subscribers for receiving calls); second, to encourage further participation through lower fixed fees (which also raises termination revenues) instead of subsidizing call receptions; third, paying customers for incoming calls reveals "extra" profits that are not used to cover deployment costs or higher long-run incremental costs, which could lead regulators to decrease the access charge.

In the United States, network operators usually offer a given amount of monthly minutes for a given monthly fee (with a per-minute rate after allowance). The three-part tariff $T(q, \widetilde{q}) = F + pq + r\widetilde{q}$ is actually equivalent to charge a fixed fee T by which customers can place and receive the volume of calls q and $\widetilde{q}$. Furthermore, network operators could price discriminate in the presence of heterogenous customers and offer different contracts $T(q, \widetilde{q})$ for each type (e.g., small and large users).

The average mobile termination rate in the European Union is currently about 8 euro cents per minute, which is still much higher than the real cost of call termination. The European Commission is planning on cutting termination rates to between 1 and 2 euro cents by 2012. Based on previous literature and the results of this paper I can draw the following implications for policy makers.

The transition (from above-cost to cost-based access charges). There are two noteworthy effects:

1. *Direct effect on participation.* The opportunity cost of servicing a customer decreases

[24] Also, in 2008 in India the mobile operator Virgin Mobile started to offer 10 paise (about 0.25 US cents) for every incoming call minute (the Department of Telecommunications of India believes that this pricing strategy is due to the termination charge of Virgin Mobile that is much higher than its cost of call termination.)

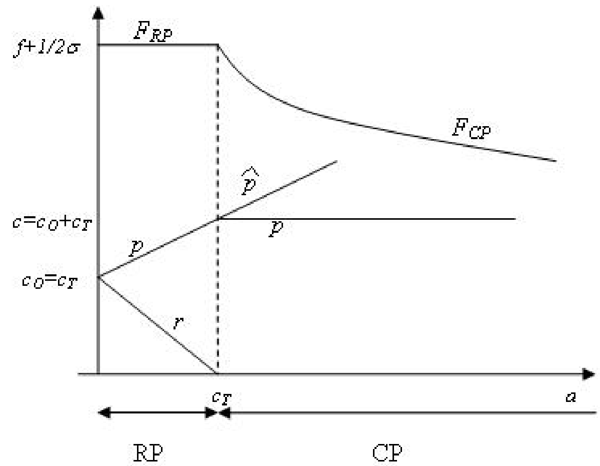

Figure 1: RP : receiver-pays regime, where $p = c + m$ is the calling price (for on- and off-net calls), $r = -m$ is the reception charge and F_{RP} is the fixed fee. CP : caller-pays regime, where $p = c$ is the on-net price, $\widehat{p} = c + m$ is the off-net price and F_{CP} is the fixed fee.

with the access charge because operators benefit from termination revenues even if a customer does not make calls. Thus, as the termination charge decreases, the opportunity cost increases. As a result, network operators will compete less aggressively for customers by charging higher fixed fees or decreasing subsidies (e.g., allowance minutes or handset subsidies), which is known as the waterbed effect. More formally, Laffont, Rey and Tirole (1998b) show that in equilibrium the fixed fees decrease with the access charge: $F_{CP}(m) = f + 1/2\sigma - v(c) + v(c + m)$, where $v(p) = \max_q\{\mu(q) - pq\}$. Gans and King (2001) show that the waterbed effect is so strong that a lower access charge could even increase the firms profit.

2. *Direct effect on consumption.* Decreasing the access charge lowers the cost of the off-net calls, which lowers the equilibrium off-net price $\widehat{p}$, a lower off-net price encourages consumption and raises the consumer surplus.

Lower fixed fees (so higher access charges) increase the participation of "small users" who are mostly interested in being called and that otherwise would not participate; their participation raises the value of the network for large users, who may call them (see Jullien and Rey, 2008 and Jullien et al. 2008). Likewise, lower off-net prices (so lower access charges) encourage consumption, which raises the utility of users. Therefore, the socially optimal access charge must consider the two effects. For instance, if 1 is dominates 2, then decreasing the access charge to the cost of an efficient operator is not necessarily socially optimal.

Cost-based access charges. If regulators set the access charge at the cost of an efficient operator, say c_T, we have $m = 0$. Then the model predicts $p = c$ and $r = 0$, i.e., efficient operators will not charge incoming calls. Inefficient operators will not recover the cost of terminating a call, which would lead them to charge customers for receiving calls. They will face a competitive disadvantage that may lead them to be efficient. Once operators charge incoming calls, regulators must be aware that termination-based price discrimination creates a connectivity-breakdown risk (see Jeon et al. 2004). So, when moving to the receiver-pays regime, termination-based price discrimination should be forbidden.

Are below-cost access charges socially optimal? If regulators find it socially optimal to decrease access charges to the cost of an efficient operator (so effect 2 dominates effect 1), then one may wonder whether it is socially optimal to decrease the access charge below cost (as in the United States). This article shows that decreasing the access charge below cost does not affect fixed fees (in a symmetric equilibrium $F = f + 1/2\sigma = F_{CP}(0)$), but it lowers the calling price and it raises the reception charge, which is then positive. Proposition 2 says that a below-cost access charge maximizes the average social welfare of a given call as long as callers and receivers have the same utility functions. Bill and keep may then be socially optimal.

Figure 1 summarizes the impact of the access charge on the equilibrium prices in the presence of the caller-pays and the receiver-pays regime.

7 Concluding remarks

This article extends the standard model of network competition by allowing mobile users to derive utility from receiving calls, receivers to affect the length of a call by hanging up, and networks to charge subscribers for making and receiving calls.

I show that network operators only charge incoming calls in the presence of below-cost access charges. I also show that when network operators find it optimal to charge users for receiving calls, then fixed fees and total discounted profits are neutral with respect to the level of the access charge even if network operators compete in a multi-period model. In equilibrium, an increase in the access charge raises the calling price and decreases the reception charge. These two effects introduce a clear distortion in the consumer welfare. The access charge that maximizes the average social welfare of a given call depends on the parameter values. I can demonstrate, however, that if the caller and the receiver have the same utility functions, then the average social welfare of such a call decreases with the access charge. Bill and keep may then be socially optimal. Finally, I discuss the implications of these results for policy makers. In particular, I discuss the impact of lower access charges, which are currently recommended by the European Commission, on the firms' strategies and social welfare.

I expect further research extending this analysis. Two key directions are noteworthy. Firstly, it is not difficult to find cases in which the calling pattern is unbalanced. For example, the existence of "closed user groups" may bias call volumes towards on-net calls, which may affect the pricing strategies. Secondly, the existence of some substitution between mobile-to-mobile and fixed-to-mobile calls may induce callers to use the cheaper alternative. Finally, this theoretical model should smooth the way for an empirical analysis that compares the social welfare in the receiver-pays regime (higher fixed fees -lower penetration rates- and lower usage prices -higher consumption) with the caller-pays regime (lower fixed fees -higher penetration rates- and higher usage prices -lower consumption).

8 APPENDIX

Proof of Proposition 1. I first focus on the network i's best response to given j's prices: p_j, r_j and F_j. For given prices p_i and r_i, $\Delta\phi \equiv \phi_i - \phi_j : [0,1] \rightarrow R$ is an affine function of the market share: $\Delta\phi(\alpha_i) = \beta + \kappa\alpha_i$, where β and κ are real numbers. Moreover, relevant fixed fees are bounded: given the pair (p_i, r_i), there exists an upper bound $\overline{F}$ so that $\overline{\alpha}_i(\overline{F}) = 0$, thus $F_i > \overline{F}$ cannot be a best response; similarly, there exists a lower bound $\underline{F}$ so that $\overline{\alpha}_i(\underline{F}) = 1$, then $F_i < \underline{F}$ cannot be a best response either because $\overline{\alpha}_i = 1$ and the i's profit decreases with F_i. Therefore, for given (p_i, r_i) the i's fixed fee that is a best response to the triple (p_j, r_j, F_j) belongs to the interval $[\underline{F}, \overline{F}]$. For given prices p_i and r_i, $\overline{\alpha}_i : [\underline{F}, \overline{F}] \longrightarrow [0,1]$ is one-to-one or injective in F_i iff $\kappa \neq 1/\sigma$: $\overline{\alpha}_i = [1/2 + (2\lambda_i - 1)\sigma s + \sigma(F_j - F_i + y)]/(1 - \sigma\kappa)$, i.e., $\overline{\alpha}_i$ is well-defined and monotonically decreasing in F_i iff $\kappa \neq 1/\sigma$. The *degenerate case* where $\kappa = 1/\sigma$ could exist for a given set of prices, however as q and $\widetilde{q}$ are bounded, I can always find a small enough σ so that this *degenerate case* cannot occur. Also, any small shock in the parameter values would make $\kappa \neq 1/\sigma$. Thus, for a small enough σ, $\overline{\alpha}_i$ is well-defined and injective, and thus invertible on its domain.

Each network i maximizes its profit π_i with respect to p_i, r_i and F_i, for given p_j, r_j and F_j, subject to $\alpha_i = (1/2) + (2\lambda_i - 1)\sigma s + \sigma(\phi_i - F_i - \phi_j + F_j)$, where π_i and ϕ_i are given by (7) and (6). From the market share definition, I can rewrite the profit as a function of p_i, r_i and α_i: $\overline{\pi}_i(p_i, r_i, \alpha_i)$, which is given by (8). $\overline{\alpha}_i$ is one-to-one in F_i, thus for given p_j, r_j and F_j, maximizing π_i with respect to p_i, r_i and F_i is equivalent to maximizing $\overline{\pi}_i$ with respect to p_i, r_i and α_i, i.e., there exists a one-to-one correspondence between the two best-response correspondences (p_i, r_i, F_i) and (p_i, r_i, α_i) to a given triple (p_j, r_j, F_j). I now show that for a small enough σ and a large enough Δ, the best-response correspondence

(p_i, r_i, α_i) is well defined, or, in other words, that the Hessian of $\overline{\pi}_i$ is negative definite: $H_i \equiv D^2\overline{\pi}_i(p_i, r_i, \alpha_i)$. Let H_i^k denote the $k-th$ principal minor of H_i. For a large enough Δ, we can write $|H_i^1| \simeq -\alpha_i/2b$, $|H_i^2| \simeq (\alpha_i)^2/4be$ and

$$\left|H_i^3\right| \simeq \lambda_{\alpha_i}\frac{(\alpha_i)^2}{2be} - \frac{1}{\sigma}\frac{(\alpha_i)^2}{2be} + \frac{\alpha_i}{2e}(\lambda_{p_i})^2 + \frac{\alpha_2^i}{2b}(\lambda_{r_i})^2,$$

where $\lambda_{p_i}(p_i, r_j, \alpha_i) \equiv [2(p_i - c)\alpha_i + ((p_i - c - m)(\alpha_j - \alpha_i) + 2\alpha_i r_j)](-1/2b)$, $\lambda_{r_i}(p_j, r_i, \alpha_i) \equiv [2\alpha_i(r_i - c) + ((\alpha_j - \alpha_i)(r_i + m) + 2p_j\alpha_i)](-1/2e)$ and $\lambda_{\alpha_i}(p_i, r_i, p_j, r_j) \equiv (1/2)(\partial^2\overline{\pi}_i/(\partial\alpha_i)^2 + 2/\sigma) = -cQ_{ii} + (c+m)Q_{ij} - mQ_{ji} + p_j(-Q_{jj} + Q_{ji}) + r_j(-Q_{jj} + Q_{ij}) + (U_{ii} + \widetilde{U}_{ii} - U_{ji} - \widetilde{U}_{ij}) - (U_{ij} + \widetilde{U}_{ji} - U_{jj} - \widetilde{U}_{jj})$.[25] Thus, for any $\alpha_i \in (0,1]$ and a large enough Δ: $|H_i^1| < 0$ and $|H_i^2| > 0$. As demands are bounded, λ_{α_i}, λ_{p_i} and λ_{r_i} are also bounded functions, then there exists a small enough σ so that $|H_i^3| < 0$. Thus, for a large enough Δ there exists a small enough σ for which the profit function is strictly concave in own strategies whatever the rival prices are, which implies that the network i's best response is a continuos function. Therefore, a candidate equilibrium must satisfy the first-order conditions (I show below that no cornered-market equilibrium exists), and an interior solution to the first-order conditions is an equilibrium. The first-order conditions can be written as follows: $(\partial\overline{\pi}_i/\partial p_i)(p_i, r_i, \alpha_i, r_j) = 0$, $(\partial\overline{\pi}_i/\partial r_i)(p_i, r_i, \alpha_i, p_j) = 0$, $(\partial\overline{\pi}_i/\partial\alpha_i)(p_i, r_i, \alpha_i, p_j, r_j, F_j) = 0$ for $i \neq j = 1, 2$. Also, the market share equation must hold. Notice that fixed fees do not enter into the first-order conditions $\partial\overline{\pi}_i/\partial p_i = 0$ and $\partial\overline{\pi}_i/\partial r_i = 0$. In particular, for $i \neq j = 1, 2$:

$$\begin{aligned} \tfrac{\partial\overline{\pi}_i}{\partial p_i} &= -\alpha_i\left(\tfrac{1}{2b}\xi_{p_i}(p_i, r_j) + \tfrac{1}{\Delta}\omega_{p_i}(p_i, r_i, r_j)\right), \\ \tfrac{\partial\overline{\pi}_i}{\partial r_i} &= -\alpha_i\left(\tfrac{1}{2e}\xi_{r_i}(r_i, p_j) + \tfrac{1}{\Delta}\omega_{r_i}(r_i, p_i, p_j)\right), \end{aligned} \tag{29}$$

where $\xi_{p_i}(p_i, r_j) = -c - \alpha_j m + p_i + \alpha_i r_j$, $\xi_{r_i}(r_i, p_j) = -\alpha_i c + \alpha_j m + r_i + \alpha_i p_j$, and $\omega_{p_i} : [\underline{r}, \overline{r}] \times [\underline{p}, \overline{p}]^2$ and $\omega_{r_i} : [\underline{r}, \overline{r}] \times [\underline{p}, \overline{p}]^2$ are nonlinear functions that do not depend on Δ.[26] So, each of these first-order conditions can be rewritten as the sum of the linear function ξ_{x_i} and the nonlinear function ω_{x_i}, where $x = p, r$ for $i = 1, 2$. This system of equations has at least one solution: $p_1 = p_2 = c + m$ and $r_1 = r_2 = -m$, where $\xi_{p_1} = \xi_{p_2} = \omega_{p_1} = \omega_{p_2} = 0$ and $\xi_{r_1} = \xi_{r_2} = \omega_{r_1} = \omega_{r_2} = 0$. Let Ξ denote the set of solutions to the system of first-order conditions, which we already know is non-empty. As Δ increases, the nonlinear components of (29) vanish. For a large enough Δ, the set Ξ is then a singleton. Therefore, for any $\alpha_i \in (0,1)$ and a large enough Δ, there exists a unique equilibrium, in this equilibrium networks charge calls at their off-net cost. Inserting $p_1 = p_2 = c + m$ and $r_1 = r_2 = -m$ into (7) yields the i's profit, which is given by (15), as a quadratic function of F_i. Hence reaction

[25] For more details see Lemma 2 of López (2007b).
[26] For more details see Lemma 1 of López (2007b).

functions $F_i^r(F_j)$, which are given by (16), are linear functions. Moreover, $dF_i^r/dF_j = 1/2$, therefore there exists a unique equilibrium in fixed fees, which is given by (17).

Finally, no cornered-market equilibrium exists. The proof of this result is an extension of the proof of Laffont et al. (1998a) to asymmetric networks. Consider a given candidate shared-market equilibrium $(p_1, r_1, F_1, p_2, r_2, F_2)$, and suppose that network 1 corners the market by charging (p_1^*, r_1^*, F_1^*). Then, $\pi_2 = 0$ and $\pi_1^* = (p_1^* - c + r_1^*)Q(p_1^*, r_1^*) + F_1^* - f$, with $\pi_1^* \geq 0$. But network 2 could charge $p_2^* = p_1^*$, $r_2^* = r_1^*$ and $F_2^* = F_1^* + \epsilon$, with $\epsilon > 0$, yielding $\alpha_2^* = (1/2) + (2(1-\alpha_0) - 1)\sigma s - \sigma\epsilon$. If $\alpha_0 = 1$, then $\alpha_2^* = (1/2)(1 - s/\tau) - \sigma\epsilon > 0$ for ϵ small enough since $s < \tau$. Therefore, there exists a small enough ϵ so that $\alpha_2^* > 0$ for any $\alpha_0 \in [0, 1]$. For such a small enough ϵ the 2's profit would be $\pi_2 = \alpha_2^*[(p_2^* - c + r_2^*)Q(p_2^*, r_2^*) + F_2^* - f] = \alpha_2^*(\pi_1^* + \epsilon) \geq \alpha_2^*\epsilon > 0$, a contradiction. ■

References

Armstrong, M., 1998. Network interconnection in telecommunications. Economic Journal 108, 545-564.

Armstrong, M., 2002. The theory of access pricing and interconnection. In: Cave, M.E., Majumdar, S.K., Vogelsang, I. (Eds.), Handbook of Telecommunications Economics. North-Holland, Amsterdam.

Armstrong, M., Wright, J., 2008. Mobile call termination. Mimeo.

Berger, U., 2004. Access charges in the presence of call externalities. Contributions to Economic Analysis & Policy 3: Iss. 1, Article 21.

Berger, U., 2005. Bill-and-Keep vs. cost-based access pricing revisited. Economic Letters 86, 107-112.

Calzada, J., Valletti,T.M., 2008. Network competition and entry deterrence. The Economic Journal 118, 1223-1244.

Carter, M., Wright, J., 2003. Asymmetric network interconnection. Review of Industrial Organization 22, 27-46.

Competition Commission., 2003. Vodafone, O2, Orange and T-Mobile: Reports on References Under Section 13 of Telecommunications Act 1984 on the Charges Made by Vodafone, O2, Orange and T-Mobile for Terminating Calls from Fixed and Mobile Networks.

De Bijl, P., Peitz, M., 2002. Regulation and entry into telecommunications markets. Cambridge, UK: Cambridge U. Press.

De Bijl, P., Peitz, M., 2004. Dynamic regulation and entry in telecommunications markets: a policy framework." Information Economics and Policy 16, 411-437.

Dessein, W., 2003. Network competition in nonlinear pricing. RAND Journal of Economics 34, 593-611.

Dessein, W., 2004. Network competition with heterogenous customers and calling patterns. Information Economics and Policy 16, 323-345.

Fabrizi, S., 2005. International telecommunications pricing: Does a scope for reform exist?. Université Toulouse 1 and Università degli Studi di Bologna.

Farrell, J., Klemperer, P., Coordination and Lock-In: Competition with Switching Costs and Network Effects. Handbook of Industrial Organization, Vol 3. M. Armstrong and R. Porter (eds.), North-Holland 2007.

Gans, J.S., King, S.P., 2000. Mobile network competition, customer ignorance and fixed-to-mobile call prices. Information Economics and Policy 12, 301-327.

Gans, J.S., King, S.P., 2001. Using 'Bill and Keep' interconnect arrangements to soften network competition. Economics Letters 71, 413-420.

Grzybowski, L., Pereira, P., 2007. Subscription choices and switching costs in mobile telephony. NET Institute Working Paper No 07-12. Available at SSRN: http://ssrn.com/abstract=1021324.

Hahn, J., 2004. Network competition and interconnection with heterogeneous subscribers. International Journal of Industrial Organization 22, 611-631.

Hermalin, B., Katz, M.L., 2001. Network interconnection with two-sided user benefits. Mimeo, University of California-Berkeley.

Hermalin, B., Katz, M.L., 2004. Sender or receiver: who should pay to exchange an electronic message?. RAND Journal of Economics 35, 423-448.

Hoernig, S., 2007. On-net and off-net pricing on asymmetric telecommunications networks. Information Economics and Policy 19:171-188.

Jeon, D.S., Hurkens, S., 2008. A retail benchmarking approach to efficient two-way access pricing: no termination-based price discrimination. RAND Journal of Economics 39, 822-849..

Jeon, D.S., Laffont, J.-J., Tirole, J., 2004. On the receiver pays principle. RAND Journal of Economics 35, 85-110.

Jullien, B., Rey, P., 2008. Notes on the economics of termination charges. IDEI Report #6 Telecommunications.

Jullien, B., Rey, P., Sand-Zantman, W., 2008. Mobile call termination revisited: Consumers heterogeneity and calling pattern. Mimeo, Toulouse School of Economics.

Kim, J.-Y., Lim. Y., 2001. An economic analysis of the receiver pays principle. Information Economics and Policy 13, 231-260.

Klemperer, P., 1987. The Competitiveness of Markets with Switching Costs. RAND Journal of Economics 18, 138-150.

Laffont, J.-J., Rey, P., Tirole, J., 1998a. Network competition: I. Overview and nondiscriminatory pricing. RAND Journal of Economics 29, 1-37.

Laffont, J.-J., Rey, P., Tirole, J., 1998b. Network competition: II. Price discrimination. RAND Journal of Economics 29, 38-56.

Laffont, J.-J., Marcus, S., Rey, P., Tirole, J., 2003. Internet interconnection and the off-net-cost pricing principle. RAND Journal of Economics 34, 370-390.

López, A.L., 2007a. Using future access charges to soften network competition. Mimeo, IESE Business School. Available at: http://angelluislopez.net.

López, A.L., 2007b. Dynamic competition in telecom networks under the receiver party pays principle. Economics of Networks -WPS: http://ssrn.com/abstract=999719.

López, A.L., Rey, P., 2008. Foreclosing competition through access charges and price discrimination. Mimeo, Toulouse School of Economics. Available at: http://idei.fr and http://angelluislopez.net.

Valletti, T.M., C. Cambini., 2005. Investments and network competition. RAND Journal of Economics 36, 446-468.